



Q257A8 A

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/37222 A1

(51) International Patent Classification⁷: G06T 17/00, G06K 9/00

(74) Agents: BERESFORD, Keith, Denis, Lewis et al.; Beresford & Co, 2-5 Warwick Court, High Holborn, London WC1R 5DJ (GB).

(21) International Application Number: PCT/GB00/04411

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(22) International Filing Date: 20 November 2000 (20.11.2000)

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9927314.6 18 November 1999 (18.11.1999) GB

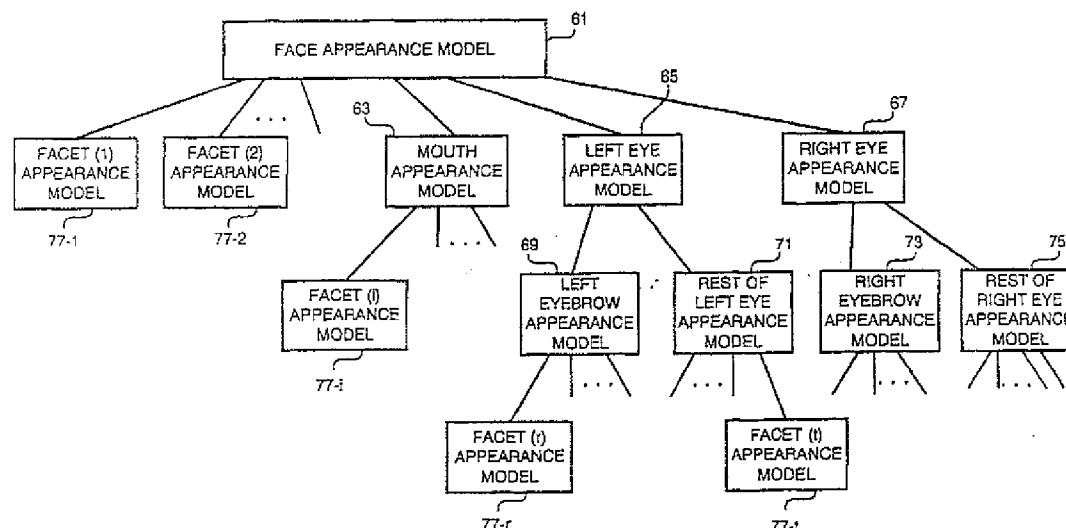
(71) Applicant (for all designated States except US): ANTHROPICS TECHNOLOGY LIMITED [GB/GB]; Ealing Studios, Ealing Green, London W5 5EP (GB).

Published:

- With international search report.
- Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMAGE PROCESSING SYSTEM



(57) Abstract: A hierarchical parametric model is provided for modelling the appearance of objects, such as human faces. The hierarchical model can model both the shape and the texture of the object. The hierarchical model includes models for components of the object being modelled such that output parameters from one model are applied as input parameters to models which are lower in the hierarchy. This hierarchical model can be used, for example, for face tracking, for video compression, for 2D and 3D character

A1

WO 01/37222

IMAGE PROCESSING SYSTEM

The present invention relates to the parametric modelling of the appearance of objects. The resulting model can be 5 used, for example, to track the object, such as a human face, in a video sequence.

The use of parametric models for image interpretation and synthesis has become increasing popular. Cootes et al 10 have shown in their paper entitled "Active Shape Models - Their Training and Application", Computer Vision and Image Understanding, Volume 61, No. 1, January, pages 38-59, 1995, how such parametric models can be used to model the variability of the shape and texture of human faces. 15 They have mainly used these models for face recognition and tracking within video sequences, although they have also demonstrated that their model can be used to model the variability of other deformable objects, such as MRI scans of knee joints. The use of these models provides 20 a basis for a broad range of applications since they explain the appearance of a given image in terms of a compact set of model parameters which can be used for higher levels of interpretation of the image. For example, when analysing face images, they can be used to 25 characterise the identity, pose or expression of a face.

Using such models for image interpretation requires, however, a method of fitting them to new image data.

This involves identifying the model parameters that generate an image which best fits (according to some measure) the new input image. Typically this problem is one of minimising the sum of squares of pixel errors

5 between the generated image and the input image. In their paper entitled "Estimating Coloured 3D Face Models from Single Images: An Example-Based Approach" Vetter and Blanz have proposed a stochastic gradient descent optimisation technique to identify the optimum model

10 parameters for the new image. Although this technique can give very accurate results finding the locally optimal solution, they generally get stuck in local minima since the error surface for the problem of fitting an appearance model to an image is particularly rough

15 containing many local minima. Therefore, this minimisation technique often fails to converge on the global minimum. An additional drawback of this technique is that it is very slow requiring several minutes to achieve convergence.

20

A faster, more robust technique known as the active appearance model was proposed by Edwards et al in the paper entitled "Interpreting Face Images using Active Appearance Models", published in the Third International Conference on Automatic Face and Gesture Recognition 1998, pages 300-305, Japan, April 1998. This technique uses a prior training stage in which the relationship between model parameter displacements and the resulting

change in image error is learnt. Although the method is much faster than direct optimisation techniques, it also requires fairly accurate initial model parameters if the search is to converge. Additionally, this technique does 5 not guarantee that the optimum parameters will be found.

The appearance model proposed by Cootes et al includes a single appearance model matrix which linearly relates a set of parameters to corresponding image data. Blanz et 10 al segmented the face into a number of completely independent appearance models, each of which is used to render a separate region of the face. The results are then merged using a general interpretation technique.

15 The present invention aims to provide an alternative way of modelling the appearance of objects which will allow subsequent image interpretation through appropriate processing of parameters generated for the image.

20 According to one aspect, the present invention provides a hierarchical parametric model for modelling the shape of an object, the model comprising data defining a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set 25 of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from one or more functions from a higher layer of the

hierarchy and to generate therefrom the relative positions of a plurality of predetermined points on the object. Such a hierarchical parametric model has the advantage that small changes in some parts of the object
5 can still be modelled by the parameters, even though they are significantly smaller than variations in other less important parts of the object. This model can be used for face tracking, video compression, 2D and 3D character generation, face recognition for security purposes, image
10 editing etc.

According to another aspect, the present invention provides an apparatus and method of determining a set of appearance parameters representative of the appearance of an object, the method comprising the steps of storing a
15 hierarchical parametric model such as the one discussed above and at least one function which relates a change in input parameters to an error between actual appearance data for the object and appearance data determined from the set of input parameters and the parametric model;
20 initially receiving a current set of input parameters for the object; determining appearance data for the object from the current set of input parameters and the stored parametric model; determining the error between the actual appearance data of the object and the appearance data determined from the current set of input parameters;
25 determining a change in the input parameters using the at least one stored function and said determined error; and

updating the current set of input parameters with the determined change in the input parameters.

5 An exemplary embodiment of the present invention will now be described with reference to the accompanying drawings in which:

10 Figure 1 is a schematic block diagram illustrating a general arrangement of a computer system which can be programmed to implement the present invention;

15 Figure 2 is a block diagram of an appearance model generation unit which receives some of the image frames of a source video sequence together with a target image frame and generates therefrom an appearance model;

20 Figure 3 is a block diagram of a target video sequence generation unit which generates a target video sequence from a source video sequence using a set of stored difference parameters;

25 Figure 4 is a flow chart illustrating the processing steps which the target video sequence generation unit shown in Figure 3 performs to generate the target video sequence;

Figure 5 schematically illustrates the form of a hierarchical appearance model generated in one embodiment

of the invention;

5 Figure 6 shows a head with a mesh of triangular facets placed over the head and whose positions are defined by the position of landmark points at the corners of the facets;

10 Figure 7 is a flow chart illustrating the processing steps required to generate a facet appearance model from the training images;

15 Figure 8 schematically illustrates the way in which a transformation is defined between a facet in a training image and a predefined shape of facet which allows texture information to be extracted from the facet;

20 Figure 9 is a flow chart illustrating the main processing steps involved in determining an appearance model for the mouth using the appearance models for the facets which appear in the mouth and using the training images;

25 Figure 10 schematically illustrates the way in which training images are used to determine some of the appearance models which form the hierarchical appearance model illustrated in Figure 5;

Figure 11a is a flow chart illustrating the processing steps performed during a training routine to identify an

Active matrix associated with a current facet;

Figure 11b is a flow chart illustrating the processing steps performed during a training routine to identify an

5 Active matrix associated with the mouth;

Figure 12 is a flow chart illustrating the processing steps involved in determining a set of parameters which define the appearance of a face within a input image;

10

Figure 13a shows three frames of an example source video sequence which is applied to the target video sequence generation unit shown in Figure 4;

15

Figure 13b shows an example target image used to generate a set of difference parameters used by the target video sequence generation unit shown in Figure 4;

20

Figure 13c shows a corresponding three frames from a target video sequence generated by the target video sequence generation unit shown in Figure 4 from the three frames of the source video sequence shown in Figure 13a using the difference parameters generated using the target image shown in Figure 13b;

25

Figure 13d shows a second example of a target image used to generate a set of difference parameters for use by the target video sequence generation unit shown in Figure 4;

and

Figure 13e shows the corresponding three frames from the target video sequence generated by the target video sequence generation unit shown in Figure 4 when the three frames of the source video sequence shown in Figure 13a are input to the target video sequence generation unit together with the difference parameters calculated using the target image shown in Figure 13d.

10

Figure 1 is an image processing apparatus according to an embodiment of the present invention. The apparatus comprises a computer 1 having a central processing unit (CPU) 3 connected to a memory 5 which is operable to store a program defining the sequence of operations of the CPU 3 and to store object and image data used in calculations by the CPU 3. Coupled to an input port of the CPU 3 there is an input device 7, which in this embodiment comprises a keyboard and a computer mouse. Instead of, or in addition to the computer mouse, another position sensitive input device (pointing device) such as a digitiser with associated stylus may be used.

25 A frame buffer 9 is also provided and is coupled to the CPU 3 and comprises a memory unit (not shown) arranged to store image data relating to at least one image, for example by providing one (or several) memory location(s) per pixel of the image. The value stored in the frame

buffer for each pixel defines the colour or intensity of that pixel in the image. In this embodiment, the images are represented by 2-D arrays of pixels, and are conveniently described in terms of Cartesian coordinates, 5 so that the position of a given pixel can be described by a pair of x-y coordinates. This representation is convenient since the image is displayed on a raster scan display 11. Therefore, the x-coordinate maps to the distance along the line of the display and the y- 10 coordinate maps to the number of the line. The frame buffer 9 has sufficient memory capacity to store at least one image. For example, for an image having a resolution of 1000 x 1000 pixels, the frame buffer 9 includes 10^6 pixel locations, each addressable directly or indirectly 15 in terms of a pixel coordinate x,y.

In this embodiment, a video tape recorder (VTR) 13 is also coupled to the frame buffer 9, for recording the image or sequence of images displayed on the display 11. 20 A mass storage device 15, such as a hard disc drive, having a high data storage capacity is also provided and coupled to the memory 5. Also coupled to the memory 5 is a floppy disc drive 17 which is operable to accept removable data storage media, such as a floppy disc 19 25 and to transfer data stored thereon to the memory 5. The memory 5 is also coupled to a printer 21 so that generated images can be output in paper form, an image input device 23 such as a scanner or video camera and a

modem 25 so that input images and output images can be received from and transmitted to remote computer terminals via a data network, such as the Internet. The CPU 3, memory 5, frame buffer 9, display unit 11 and mass storage device 13 may be commercially available as a complete system, for example as an IBM compatible personal computer (PC) or a workstation such as the Sparc station available from Sun Microsystems.

10 A number of embodiments of the invention can be supplied commercially in the form of programs stored on a floppy disc 19 or on other mediums, or as signals transmitted over a data link, such as the Internet, so that the receiving hardware becomes reconfigured into an apparatus 15 embodying the present invention.

20 In this embodiment, the computer 1 is programmed to receive a source video sequence input by the image input device 23 and to generate a target video sequence from the source video sequence using a target image. In this embodiment, the source video sequence is a video clip of an actor acting out a scene, the target image is an image of a second actor and the resulting target video sequence is a video sequence showing the second actor acting out 25 the scene. The way in which this is achieved will now be briefly described with reference to Figures 2 to 4.

In this embodiment, in order to generate the target video

sequence from the source video sequence, a hierarchical parametric appearance model which models the variability of shape and texture of the head images is used. This appearance model makes use of the fact that some prior 5 knowledge is available about the contents of head images in order to facilitate their modelling. For example, it can be assumed that two frontal images of a human face will each include eyes, a nose and a mouth. In this embodiment, as shown in Figure 2, the hierarchical parametric appearance model 35 is generated by an 10 appearance model generation unit 31 from training images which are stored in an image database 32. In this embodiment, all the training images are colour images having 500 x 500 pixels, with each pixel having a red, 15 green and a blue pixel value. The resulting appearance model 35 is a parameterisation of the appearance of the class of head images defined by the heads in the training images, so that a relatively small number of parameters (for example 50) can describe the detailed (pixel level) 20 appearance of a head image from the class. In particular, the hierarchical appearance model 35 defines a function (F) such that:

$$I = F(p) \quad (1)$$

where p is the set of appearance parameters (written in vector notation) which generates, through the hierarchical appearance model (F), the face image I . The

structure of the hierarchical appearance model used in this embodiment will be described later.

Once the hierarchical appearance model 35 has been 5 determined, a target video sequence can be generated from a source video sequence. As shown in Figure 3, the source video sequence is input to a target video sequence generation unit 51 which processes the source video sequence using a set of difference parameters 53 to generate and to output the target video sequence. The difference parameters 53 are determined by subtracting the appearance parameters which are generated for the first actor's head in one of the source video frames, from the appearance parameters which are generated for 10 the second actor's head in the target image. The way in which these appearance parameters are determined for these images will be described later. In order that these difference parameters only represent differences in the general shape and colour texture of the two actors' heads, the pose and facial expression of the first actor's head in the source video frame used should match, as closely as possible, the pose and facial expression of 15 the second actor's head in the target image.

20 The processing steps required to generate the target video sequence from the source video sequence will now be described in more detail with reference to Figure 4. As shown, in step s1, the appearance parameters (p_s^i) for 25

the first actor's head in the current video frame (I_{s^i}) are automatically calculated. The way that this is achieved will be described later. Then, in step s3, the difference parameters (p_{dif}) are added to the appearance parameters for the first actor's head in the current video frame to generate:

$$p_{mod}^i = p_s^i + p_{dif} \quad (2)$$

10 The resulting appearance parameters (p_{mod}^i) are then used, in step s5, to regenerate the head for the current target video frame. In particular, the modified appearance parameters are inserted into equation (1) above to regenerate a modified head image which is then 15 composited, in step s7, into the source video frame to generate the corresponding target video frame. A check is then made, in step s9, to determine whether or not there are any more source video frames. If there are, then the processing returns to step s1 where the 20 procedure described above is repeated for the next source video frame. If there are no more source video frames, then the processing ends.

25 Figure 13 illustrates the results of this animation technique (although showing black and white images and not colour). In particular, Figure 13a shows three frames of the source video sequence, Figure 13b shows the target image (which in this embodiment is computer

generated) and Figure 13c shows the corresponding three frames of the target video sequence obtained in the manner described above. As can be seen, an animated sequence of the computer generated character has been 5 generated from a video clip of a real person and a single image of the computer generated character.

HIERARCHICAL APPEARANCE MODEL

In the systems described by Cootes et al and Blanz et al, 10 the parametric model is created by placing a number of landmark points on a training image and then identifying the same landmark points on the other training images in order to identify how the location of and the pixel values around the landmark points vary within the 15 training images. A principal component analysis is then performed on the matrix which consists of vectors of the landmark points. This PCA yields a set of Eigenvectors which describe the directions of greatest variation along which the landmark points change. Their appearance model 20 includes the linear combination of the Eigenvectors plus parameters for translation, rotation and scaling. This single appearance model relates a compact set of appearance parameters to pixel values.

25 In this embodiment, rather than having a single appearance model for the object, a hierarchical appearance model comprising several appearance models which model variations in components of the object is

used. For example, in the case of human faces, the hierarchical appearance model may include an appearance model for the mouth, one for the left eye, one for the right eye and one for the nose. Since it may be possible 5 to model various components of the object, the particular hierarchical structure which will be used for a particular object and application must first of all be defined by the system designer.

10 Figure 5 schematically illustrates the structure of the hierarchical appearance model used in this embodiment. As shown, at the top of the hierarchy there is a general face appearance model 61. Beneath the face appearance model there is a mouth appearance model 63, a left eye appearance model 65, a right eye appearance model 67, a left eyebrow appearance model 69, a rest of left eye appearance model 71, a right eyebrow appearance model 73, a rest of right eye appearance model 75 and, in this 15 embodiment, a facet appearance model for each facet defined in the training images. Figure 6 shows the head of a training image in which the set of landmark points 20 has been placed at the appropriate points on the head. As shown, in this embodiment, there are one hundred and forty-eight triangular areas or facets defined by the 25 positions of the landmark points. Therefore, in this embodiment, there are one hundred and forty-eight facet appearance models 77.

The face appearance model 61 operates to relate a small number of "global" appearance parameters to a further set of appearance parameters, some of which are input to facet appearance models 77, some of which are input to the mouth appearance model 63, some of which are input to the left eye appearance model 65 and the rest of which are input to the right eye appearance model 67. The facet appearance models 77 operate to relate the input parameters received from the appearance model which is above it in the hierarchy into corresponding pixel values for that facet. The mouth appearance model 63 is operable to relate the parameters it receives from the face appearance model 61 into a further set of appearance parameters, respective ones of which are output to the respective facet appearance models 77 for the facets which are associated with the mouth. Similarly, the left and right eye appearance models 65 and 67 operate to relate the parameters it receives from the face appearance model 61 into a further set of appearance parameters, some of which are input to the appropriate eyebrow appearance model and the rest of which are input to the appropriate rest of eye appearance model. These appearance models in turn convert these parameters into parameters for input to the facet appearance models associated with the facets which appear in the left and right eyes respectively. In this way, a small compact set of "global" appearance parameters input to the face appearance model 61 can filter through the hierarchical

structure illustrated in Figure 5 to generate a set of pixel values for all the facets in a head which can then be used to regenerate the image of the head.

5 The way in which the individual appearance models of this hierarchical appearance model are generated in this embodiment will now be described with reference to Figures 6 to 10.

10 In this embodiment, each of the training images stored in the image database 32 is labelled with eighty six landmark points. In this embodiment, this is performed manually by the user via the user interface 33. In particular, each training image is displayed on the display 11 and the user places the landmark points over the head in the training image. These points delineate the main features in the head, such as the position of the hairline, neck, eyes, nose, ears and mouth. In order to compare training faces, each landmark point is 15 associated with the same point on each face. In this 20 embodiment, the following landmark points are used:

Landmark Point	Associated Position	Landmark Point	Associated Position
LP ₁	Left corner of left eye	LP ₄₄	Eye, bottom
LP ₂	Right corner of right eye	LP ₄₅	Eye, top
LP ₃	Chin, bottom	LP ₄₆	Eye, bottom
LP ₄	Right corner of left eye	LP ₄₇	Eyebrow, lower

Landmark Point	Associated Position	Landmark Point	Associated Position
LP ₅	Left corner of right eye	LP ₄₈	Eyebrow, upper
LP ₆	Mouth, left	LP ₄₉	Cheek, left
LP ₇	Mouth, right	LP ₅₀	Cheek, right
LP ₈	Nose, bottom	LP ₅₁	Eyebrow, lower
5	LP ₉	LP ₅₂	Eyebrow, upper
	LP ₁₀	LP ₅₃	Eyebrow, lower
	LP ₁₁	LP ₅₄	Eyebrow, upper
	LP ₁₂	LP ₅₅	Eyebrow, lower
	LP ₁₃	LP ₅₆	Eyebrow, upper
	LP ₁₄	LP ₅₇	Eyebrow, lower
	LP ₁₅	LP ₅₈	Eyebrow, upper
	LP ₁₆	LP ₅₉	Eyebrow, lower
	LP ₁₇	LP ₆₀	Eyebrow, upper
	LP ₁₈	LP ₆₁	Eyebrow, lower
10	LP ₁₉	LP ₆₂	Lower lip, top
	LP ₂₀	LP ₆₃	Centre forehead
	LP ₂₁	LP ₆₄	Upper lip, top left
	LP ₂₂	LP ₆₅	Upper lip, top right
	LP ₂₃	LP ₆₆	Lower lip, bottom right
	LP ₂₄	LP ₆₇	Lower lip, bottom left
15	LP ₂₅	LP ₆₈	Eye, top left
	LP ₂₆	LP ₆₉	Eye, top right
	LP ₂₇	LP ₇₀	Eye, bottom right
	LP ₂₈	LP ₇₁	Eye, bottom left
	LP ₂₉	LP ₇₂	Eye, top left
	LP ₃₀	LP ₇₃	Eye, top right
20	LP ₃₁	LP ₇₄	Eye, bottom right

Landmark Point	Associated Position	Landmark Point	Associated Position
LP ₃₂	Shoulder	LP ₇₃	Eye, bottom left
LP ₃₃	Bottom, left	LP ₇₆	Lower lip, top left
LP ₃₄	Bottom, middle	LP ₇₇	Lower lip, top right
LP ₃₅	Bottom, right	LP ₇₈	Chin, left
5 LP ₃₆	Left forehead	LP ₇₉	Chin, right
LP ₃₇	Right forehead	LP ₈₀	Neck, left
LP ₃₈	Centre, between eyebrows	LP ₈₁	Neckline, left
LP ₃₉	Nose, left	LP ₈₂	Neckline
LP ₄₀	Nose, right	LP ₈₃	Neckline, right
10 LP ₄₁	Nose edge, left	LP ₈₄	Neck, right
LP ₄₂	Nose edge, right	LP ₈₅	Hair edge
LP ₄₃	Eye, top	LP ₈₆	Hair edge

The result of the manual placement of the landmark points
15 is a table of landmark points for each training image,
which identifies the (x, y) coordinate of each landmark
point within the image. As shown in Figure 6, these
landmark points are also used to define the location of
predetermined triangular facets or areas within the
20 training image.

FACET APPEARANCE MODEL

Figure 7 shows a flow chart illustrating the main processing steps involved in this embodiment in
25 determining a facet appearance model for facet (i). As shown, in step s61, the system determines, for each training image, the apex coordinates of facet (i) and

texture values from within facet (i). In order to sample texture from within the facet at corresponding points within each training facet, a transformation which transforms the facet onto a reference facet is determined. Figure 8 illustrates this transformation. In particular, Figure 8 shows facet f_i^v taken from the v -th training image, which is defined by the landmark points (x_1^v, y_1^v) , (x_2^v, y_2^v) and (x_3^v, y_3^v) . The transformation (T_i^v) which transforms those coordinates onto coordinates $(0,0)$, $(1,0)$ and $(0,1)$ is determined. In this embodiment, the texture information extracted from each training facet is defined by the regular array of pixels shown in the reference facet. In order to determine the corresponding red, green and blue pixel values in the training image, the inverse transformation $([T_i^v]^{-1})$ is used to transform the pixel locations in the reference facet, into corresponding locations in the training facet, from which the RGB pixel values are determined. In this embodiment, this transformation may not result in an exact correspondence with a single image pixel location since the pixel resolution in the actual facet may be different to the resolution in the reference facet. In this embodiment, the texture information (RGB pixel values) which is determined is obtained by interpolating between the surrounding image RGB pixel values. In this embodiment, there are fifty pixels in the regular array of pixels in the reference facet. Therefore, fifty RGB pixel values are extracted for each

training facet. The texture information for facet (i) from the V-th training image can then be represented by a vector (t^{iv}) of the form:

5
$$\underline{t}^{iv} = [t_1^{iv}, t_2^{iv}, t_3^{iv} \dots t_{50}^{iv}]^T$$

where t_1^{iv} is the RGB texture information for the first reference pixel extracted from facet (i) in the V-th training image etc.

10

In this embodiment, the facet appearance models 77 treat shape and texture separately. Therefore, in step s63, the system performs a principal component analysis (PCA) on the set of texture training vectors generated in step 15 s61. For a more detailed discussion of principal component analysis, the reader is referred to the book by W. J. Krzanowski entitled "Principles of Multivariate Analysis - A User's Perspective" 1998, Oxford Statistical Science Series. As those skilled in the art will appreciate, this principal component analysis determines all possible modes of variation within the training texture vectors. However, since each of the facets is associated with a similar point on the face, most of the variation within the data can be explained by a few modes 20 of variation. The result of the principal component analysis is a facet texture appearance model (defined by matrix F_i) which relates a vector of facet texture 25 parameters to a vector of texture pixel values, by:

$$p_v^{Fit} = F_i (\underline{t}^{iv} - \underline{\bar{t}}^i) \quad (3)$$

where \underline{t}^{iv} is the RGB texture vector defined above, $\underline{\bar{t}}^i$ is the mean RGB texture vector for facet (i), F_i is a matrix which defines the facet texture appearance model for facet (i) and p^{Fit}_v is a vector of the facet texture parameters which describes the RGB texture vector \underline{t}^{iv} . The matrix F_i describes the main modes of variation of the texture within the training facets; and the vector of facet texture parameters (p^{Fit}_v) for a given input facet has a parameter associated with each mode of variation whose value relates the texture of the input facet to the corresponding mode of variation.

As those skilled in the art will appreciate, for facets which describe fairly constant parts of the face, such as the chin or cheeks, very few parameters will be needed to model the variability within the training images. However, facets which are associated with areas of the face where there is a large amount of variability (such as facets which form part of the eye), will require a larger number of facet texture parameters to describe the variability within the training images. Therefore, in step s65, the system determines how many texture parameters are needed for the current facet and stores the appropriate facet appearance model matrix.

parameters p^{fit}_v for a given texture vector \underline{t}^{iv} , equation (3) can be solved with respect to the texture vector \underline{t}^{iv} to give:

5

$$\underline{t}^{iv} = \underline{t}^i - F_i^T D_v^{fit} \quad (4)$$

10

since $F_i F_i^T$ equals the identity matrix. Therefore, by modifying the set of texture parameters (p^{fit}) within suitable limits, new textures for facet (i) can be generated which are similar to those in the training set.

15

20

Once the above procedure has been performed for each of the one hundred and forty-eight facets in the training images, a facet texture appearance model will have been generated for each of those facets. In this embodiment, the facet appearance model does not compress the parameters defining the shape of the facets, since only six parameters are needed to define the shape of each facet - two parameters for each (x,y) coordinate of the facet's apexes.

MOUTH APPEARANCE MODEL

Figure 9 shows a flow chart illustrating the main processing steps required in order to generate the mouth appearance model 63. As shown, in step s67, the system uses the facet appearance models for the facets which form part of the mouth to generate shape and texture parameters from those facets for each training image.

Therefore, referring to Figure 10, the mouth appearance model 63 will receive texture and shape parameters from the facet appearance model for facet (i), facet (j) and facet (n) for the corresponding facets in each of the 5 training images 79. As illustrated in Figure 10, the appearance model for facet (i) is operable to generate, for each training image, six shape parameters (corresponding to the three (x,y) coordinates of the apexes of facet (i)) and six texture parameters. 10 Similarly, the appearance model for facet (j) is operable to generate, for each training image, six shape parameters and four texture parameters and the appearance model for facet (n) is operable to generate, for each training image, six shape parameters and three texture 15 parameters.

The processing then proceeds to step s69 where the system performs a principal component analysis on the shape and texture parameters generated for the training images by the facet appearance models associated with the mouth. 20 In this embodiment, the mouth appearance model 63 treats the shape and texture separately. In particular, for each training image, the system concatenates the six shape parameters for the facets associated with the mouth to form the following shape vector:

$$P^{MS} = [x_1^{fi}, y_1^{fi}, x_2^{fi}, y_2^{fi}, x_3^{fi}, y_3^{fi} : x_1^{fj}, y_1^{fj}, x_2^{fj}, y_2^{fj} \dots]^T$$

and concatenates the facet texture parameters output by the facet appearance models associated with the mouth to form the following texture vector:

5 $\underline{p}^{Mt} = [p_1^{Fit}, p_2^{Fit} \dots p_6^{Fit} : p_1^{Fjt}, p_2^{Fjt} \dots p_4^{Fjt} : p_1^{Fnt}, \dots]^T$

10 The system then performs a principal component analysis on the shape vectors generated by all the training images to generate a shape appearance model for the mouth (defined by matrix M_s) which relates each mouth shape vector to a corresponding vector of shape mouth parameters by:

15
$$\underline{p}_v^{Ms} = M_s (\underline{p}_v^{Ms} - \bar{\underline{p}}^{Ms}) \quad (5)$$

20 where \underline{p}_v^{Ms} is the mouth shape vector for the mouth in the v -th training image, $\bar{\underline{p}}^{Ms}$ is the mean mouth shape vector from the training vectors and \underline{p}_v^{Ms} is a vector of mouth shape parameters for the mouth shape vector \underline{p}_v^{Ms} . The mouth shape model, defined by matrix M_s , describes the main modes of variation of the shape of the mouths within the training images; and the vector of mouth shape parameters (\underline{p}_v^{Ms}) for the mouth in the v -th training image has a parameter associated with each mode of variation whose value relates the shape of the input mouth to the corresponding mode of variation.

25 As with the facet appearance models, equation (5) above

can be rewritten with respect to the mouth shape vector p^{MS}_v to give:

$$p^{MS}_v = \bar{p}^{MS} - M_s^T \bar{p}^{MS}_v \quad (6)$$

5

since $M_s M_s^T$ equals the identity matrix. Therefore, by modifying the mouth shape parameters, new mouth shapes can be generated which will be similar to those in the training set.

10

The system then performs a principal component analysis on the mouth texture parameter vectors (p^{Mt}) which are generated for the training images. This principal component analysis generates a mouth texture model 15 (defined by matrix M_t) which relates each of the facet texture parameter vectors for the facets associated with the mouth, to a corresponding vector of mouth texture parameters, by:

20

$$\bar{p}^{Mt}_v = M_t (p^{Mt}_v - \bar{p}^{Mt}) \quad (7)$$

25

where p^{Mt}_v is a vector of mouth facet texture parameters generated by the facet appearance models associated with the mouth for the mouth in the v -th training image; \bar{p}^{Mt} is the mean vector of mouth facet texture parameters from the training vectors and p^{Mt}_v is a vector of mouth texture parameters for the facet texture parameters p^{Mt}_v . The matrix M_t describes the main modes of variation within

the training images of the facet texture parameters generated by the facet appearance models which are associated with the mouth; and the vector of mouth texture parameters (p^{Mt}_v) has a parameter associated with each of those modes of variation whose value relates the texture of the input mouth to the corresponding mode of variation.

5 The processing then proceeds to step s71 shown in Figure 10 9 where the system determines the number of shape parameters and texture parameters needed to describe the training data received from the facet appearance models which are associated with the mouth. As shown in Figure 10, in this embodiment, the mouth appearance model 63 15 requires five shape parameters and four texture parameters to be able to model most of this variation. The system therefore stores the appropriate mouth shape and texture appearance model matrices for subsequent use.

20 As those skilled in the art will appreciate, a similar procedure is performed to determine each of the appearance models shown in Figure 5, starting from the facet appearance models at the base of the hierarchy. A further description of how these remaining appearance 25 models are determined will, therefore, not be given here. The resulting hierarchical appearance model allows a small number of global face appearance parameters to be input to the face appearance model 61, which generates

further parameters which propagate down through the hierarchical model structure until facet pixel values are generated, from which an image which corresponds to the global appearance parameters can be generated.

5

AUTOMATIC GENERATION OF APPEARANCE PARAMETERS

In the description given above of the way in which the appearance models are generated, appearance parameters for an image were generated from a manual placement of a 10 number of landmark points over the image. However, during use of the appearance model to track the first actor's head in the source video sequence and during the calculation of the difference parameters (p_{diff}), the appearance parameters for the heads in the input images 15 were automatically calculated. This task involves finding the set of global appearance parameters p which best describe the pixels in view. This problem is complicated because the inverse of each of the appearance models in the hierarchical appearance model is not 20 necessarily one-to-one. In this embodiment, the appearance parameters for the head in an input image are calculated in a two-step process. In the first step, an initial set of global appearance parameters for the head in the current frame (I_s^i) is found using a simple and 25 rapid technique. For all but the first frame of the source video sequence, this is achieved by simply using the appearance parameters from the preceding video frame (I_s^{i-1}) before modification in step s3 (i.e. parameters

p_s^{i-1}). In this embodiment, the global appearance parameters (p) effectively define the shape and colour texture of the head. For the first frame and for the target image the initial estimate of the appearance parameters is set to the mean set of appearance parameters and the scale, position and orientation is initially estimated by the user manually placing the mean head over the head in the image.

In the second step, an iterative technique is used in order to make fine adjustments to the initial estimate of the appearance parameters. The adjustments are made in an attempt to minimise the difference between the head described by the global appearance parameters (the model head) and the head in the current video frame (the image head). With 50 appearance parameters, this represents a difficult optimisation problem. This can be performed by using a standard steepest descent optimisation technique to iteratively reduce the mean squared error between the given image pixels and those predicted by a particular set of appearance parameter values. In particular, minimising the following error function $E(p)$:

$$E(p) = [I^a - F(p)]^T [I^a - F(p)] \quad (8)$$

where I^a is a vector of actual image RGB pixel values at the locations where the appearance model predicts values

(the appearance model does not predict all pixel values since it ignores background pixels and only predicts a subsample of pixel values within the object being modelled) and $F(p)$ is the vector of image RGB pixel values predicted by the hierarchical appearance model. 5 As those skilled in the art will appreciate, $E(p)$ will only be zero when the model head (i.e. $F(p)$) predicts the actual image head (I^a) exactly. Standard steepest descent optimisation techniques stipulate that a step in the direction $-\nabla E(p)$ should result in a reduction in the error function $E(p)$, provided the error function is well behaved. 10 Therefore, the change (Δp) in the set of parameter values should be:

$$15 \quad \Delta \bar{p} = 2 [\nabla F(p)]^T [I^a - F(p)] \quad (9)$$

which requires the calculation of the differential of the appearance model, i.e. $\nabla F(p)$.

20 The technique described by Edwards et al assumes that, on average over the whole parameter space, $\nabla F(p)$ is constant. The update equation then becomes:

$$25 \quad \Delta p = A [I^a - F(p)] \quad (10)$$

for some constant matrix A (referred to as the "Active matrix") which is determined beforehand during a training routine. In this embodiment, rather than using a single

constant matrix associated with the entire hierarchical appearance model, an Active matrix is determined and used for each of the individual appearance models which form part of the hierarchical appearance model. The way in 5 which these Active matrices are determined in this embodiment will now be described with reference to Figures 11a and 11b, which illustrate the processing steps performed to generate the Active matrix for each facet appearance model and the Active matrix for the 10 mouth appearance model.

As shown in Figure 11a, in step s73, the system chooses a random facet parameter vector (p^{fi}) for the current facet (i) and then, in s75, perturbs this facet parameter vector by a small random amount to create $p^{fi} + \Delta p^{fi}$. In 15 this embodiment, the facet parameter vectors include not only the texture parameters, but also the six shape parameters which define the (x,y) coordinates of the facet's location within the image. The processing then proceeds to step s77 where the system uses the parameter 20 vector p^{fi} and the perturbed parameter vector $p^{fi} + \Delta p^{fi}$ to create model images I_0^{fi} and I_1^{fi} respectively. The processing then proceeds to step s79 where the system records the parameter change Δp^{fi} and image difference $I_1^{fi} - I_0^{fi}$. Then in step s81, the system determines whether 25 or not there is sufficient training data for the current facet. If there is not then the processing returns to step s21. Once sufficient training data has been

generated, the processing proceeds to step s83 where the system performs multiple multivariate linear regressions on the data for the current facet to identify an Active matrix (A_{pi}) for the current facet.

5

Figure 11b shows the processing steps required to calculate the Active matrix for the mouth appearance model. As shown, in step s85, the system chooses a random mouth parameter vector p^* . In this embodiment, this vector includes both the mouth shape parameters and the mouth texture parameters. Then, in step s87, the system perturbs this mouth parameter vector by a small random amount to create $p^* + \Delta p^*$. The processing then proceeds to step s89 where the system uses the mouth parameter vectors p^* and the perturbed mouth parameter vector $p^* + \Delta p^*$ to create model images I_0^* and I_1^* respectively, using the mouth appearance model and the facet appearance models associated with the mouth. The processing then proceeds to step s91 where the facet appearance models associated with the mouth are used again to transform the mouth model images I_0^* and I_1^* into corresponding facet appearance parameters p_0^{FM} and p_1^{FM} , which are then subtracted to determine the corresponding change Δp^{FM} in the mouth facet parameters. The processing then proceeds to step s93 where the system records the mouth parameter change Δp^* and the mouth facet parameter change Δp^{FM} . The processing then proceeds to step s95 where the system determines whether

or not there is sufficient training data. If there is not, then the processing returns to step s85. Once sufficient training data has been generated, the processing proceeds to step s97, where the system 5 performs multiple multivariate linear regressions on the training data for the mouth to identify the Active matrix (A_M) for the mouth which relates changes in mouth parameters Δp^M to changes in facet parameters Δp^{FM} for the facets associated with the mouth.

10

As those skilled in the art will appreciate, a similar processing technique is used in order to identify the Active matrix for each of the appearance models shown in Figure 5.

15

Once the Active matrices have been determined for the hierarchical appearance model, they can then be used to iteratively update a current estimate of a set of appearance parameters for an input image. Figure 12 20 illustrates the processing steps performed in this iterative routine for the current source video frame. As shown, in step s101, the system initially estimates a set of global parameters for the head in the current source video frame. The processing then proceeds to step s103 where the system generates a model image from the estimated global parameters and the hierarchical appearance model. The system then proceeds to step s105 25 where it determines the image error between the model

image and the current source video frame. Then, in step s107, the system uses this image error to propagate parameter changes up the hierarchy of the hierarchical appearance model using the stored Active matrices to 5 determine a change in the global parameters. This change in global parameters is then used, in step s109, to update the current global parameters for the current source video frame. The system then determines, in step s111, whether or not convergence has been reached by 10 comparing the error obtained from equation (8) using the updated global parameters with a predetermined threshold (Th). If convergence has not been reached, then the processing returns to step s103. Once convergence is reached, the processing proceeds to step s113, where the 15 current global appearance parameters are output as the global appearance parameters for the current source video frame and then the processing ends.

ALTERNATIVE EMBODIMENTS

20 In the above embodiment, the same hierarchical model structure was used to model the variation in the shape and texture within the training images. As those skilled in the art will appreciate, one model hierarchy can be used to model the shape variation and a different model hierarchy can be used to model the texture variation. 25 Alternatively still, rather than separating the shape and texture parameters, each of the appearance models within the hierarchical model may model the combined variation

of the shape and texture within the training images.

In the above embodiments, a facet appearance model was generated for each facet defined within the training 5 images. As those skilled in the art will appreciate, many of the facets may be grouped together such that a single facet appearance model is generated for those facets. In one form of such an embodiment, a single facet appearance model may be determined which models the 10 variability of texture within each facet of the training images.

In the above embodiments, the same amount of texture information was extracted from each facet within the 15 training images. In particular, fifty RGB texture values were extracted from each training facet. In an alternative embodiment, the amount of texture information extracted from each facet may vary in dependence upon the size of the facet. For example, more texture information 20 may be extracted from larger facets or more texture information may be extracted from facets associated with important features of the face, such as the mouth, eyes or nose.

25 In the above embodiments, each appearance model was determined from a principal component analysis of a set of training data. This principal component analysis determines a linear relationship between the training

data and a set of model parameters. As those skilled in the art will appreciate, techniques other than principal component analysis can be used to determine a parametric model which relates a set of parameters to the training data. 5 This model may define a non-linear relationship between the training data and the model parameters. For example, one or more of the models within the hierarchy may comprise a neural network which relates the set of input parameters to the training data.

10

In the above embodiments, a principal component analysis was performed on a set of training data in order to identify a relatively small number of parameters which describe the main modes of variation within the training data. 15 This allows a relatively small number of input parameters to be able to generate a larger set of output parameters from the model. However, as those skilled in the art will appreciate, this is not essential. One or more of the appearance models may act as transformation models in which the number input parameters is the same as or greater than the number of output parameters. This can be used to generate a set of input parameters which can be changed by the user in some intuitive way. For example, in order to identify parameters which have a 20 linear relationship with features in the object, such as a parameter that linearly changes the amount of smile within a face image.

25

In the above embodiments, a set of Active matrices were used in order to identify automatically a set of appearance parameters for an input image. As those skilled in the art will appreciate, rather than having 5 separate Active matrices for each of the components in the hierarchical appearance model, a global Active matrix may be used instead. Further, although both the shape and grey level parameters were used in order to derive the Active matrices, suitable Active matrices can be 10 determined using just the shape information.

In the above embodiments, the variation in both the shape and texture within the training images were modelled. As those skilled in the art will appreciate, this 15 hierarchical modelling technique can be used to model only the shape of the objects within the training images. Such a shape model could be then used to track objects within a video sequence.

20 In the first embodiment, the target image illustrated a computer generated head. This is not essential. For example, the target image might be a hand-drawn head or an image of a real person. Figures 13d and 13e illustrate how an embodiment with a hand-drawn character 25 might be used in character animation. In particular, Figure 13d shows a hand-drawn sketch of a character which, when combined with the images from the source video sequence (some of which are shown in Figure 13a)

generate a target video sequence, some frames of which are shown in Figure 13e. As can be seen from a comparison of the corresponding frames in the source and target video frames, the hand-drawn sketch has been 5 animated automatically using this technique. As those skilled in the art will appreciate, this is a much quicker and simpler technique for achieving computer animation as compared with existing systems which require the animator to manually create each frame of the 10 animation. In particular, in this embodiment, all that is required is a video sequence of a real life actor acting out the scene to be animated, together with a single sketch of the character to be animated.

15 The above embodiment has described the way in which a target image can be used to modify a source video sequence. In order to do this, a set of appearance parameters has to be automatically calculated for each frame in the video sequence. This involved the use of a 20 number of Active matrices which relate image errors to appearance parameter changes. As those skilled in the art will appreciate, similar processing is required in other applications, such as the tracking of an object within a video sequence, the tracking of a human face 25 within a video sequence or the tracking of a knee joint in an MRI scan.

In the above embodiment, the appearance model was used to

model the variations in facial expressions and 3D pose of human heads. As those skilled in the art will appreciate, the appearance model can be used to model the appearance of any deformable object such as parts of the body and other animals and objects. For example, the above techniques can be used to track the movement of lips in a video sequence. Such an embodiment could be used in film dubbing applications in order to synchronise the lip movements with the dubbed sound. This animation technique might also be used to give animals and other objects human-like characteristics by combining images of them with a video sequence of an actor. This technique can also be used for monitoring the shape and appearance of objects passing along a production line for quality control purposes.

In the above embodiment, the appearance model was generated by using a principal component analysis of shape and texture data which is extracted from the training images. As those skilled in the art will appreciate, by modelling the features of the training heads in this way, it is possible to accurately model each head by just a small number of parameters. However, other modelling techniques, such as vector quantisation and wavelet techniques can be used.

In the above embodiments, the training images used to generate the appearance model were all colour images in

which each pixel had an RGB value. As those skilled in the art will appreciate, the way in which the colour is represented in this embodiment is not important. In particular, rather than each pixel having a red, green and blue value, they might be represented by a chrominance and a luminance component or by hue, saturation and value components. Alternatively still, the training images may be black and white images, in which case only grey level data would be extracted from the facets in the training images. Additionally, the resolution of each training image may be different.

In the above embodiment, during the automatic generation of the appearance parameters, and in particular during the iterative updating of these appearance parameters the error between the input image and the model image was generated using the appearance model. Since this iterative technique still requires a relatively accurate initial estimate for the appearance parameters, it is possible initially to perform the iterations using lower resolution images and once convergence has been reached for the lower resolutions to then increase the resolution of the images and to repeat the iterations for the higher resolutions. In such an embodiment, separate Active matrices would be required for each of the resolutions.

In the above embodiment, the difference parameters were determined by comparing the image of the first actor from

one of the frames of the source video sequence with the image of the second actor in the target image. In an alternative embodiment, a separate image of the first actor may be provided which does not form part of the 5 source video sequence.

In the above embodiments, each of the appearance models modelled variations in two-dimensional images. The above modelling technique could be adapted to work with 3D 10 images and animations. In such an embodiment, the training images used to generate the appearance model would normally include 3D images instead of 2D images. The three-dimensional models may be obtained using a three dimensional scanner which typically work either by 15 using laser range-finding over the object or by using one or more stereo pairs of cameras. Once a 3D hierarchical appearance model has been created from the training models, new 3D models can be generated by adjusting the appearance parameters and existing 3D models can be 20 animated using the same differencing technique that was used in the two-dimensional embodiment described above. This 3D model can then be used to track 3D objects directly within a 3D animation. Alternatively, a 2D model may be used to track the 3D object within a video 25 sequence and then use the result to generate 3D data for the tracked object.

In the above embodiment, a set of difference parameters

were identified which describe the main differences between the head in the video sequence and the head in the target image, which difference parameters were used to modify the video sequence so as to generate a target 5 video sequence showing the second head. In the embodiment, the set of difference parameters were added to a set of appearance parameters for the current frame being processed. In an alternative embodiment, the difference parameters may be weighted so that, for 10 example, the target video sequence shows a head having characteristics from both the first and second actors.

In the above embodiment, a hierarchical appearance model 15 is used to model the appearance of human faces. The model is then used to modify a source video sequence showing a first actor performing a scene to generate a target video sequence showing a second actor performing the same scene. As those skilled in the art will appreciate, the hierarchical model presented above can be 20 used in various other applications. For example, the hierarchical appearance model can be used for synthetic two-dimensional or three-dimensional character generation; video compression when the video is substantially that of an object which is modelled by the 25 appearance model; object recognition for security purposes; face tracking for human performance analysis or human computer interaction and the like; 3D model generation from two-dimensional images; and image editing

(for example making people look older or younger, fatter or thinner etc).

In the above embodiment, an iterative process was used to
5 update an estimated set of appearance parameters for an
input image. This iterative process continued until an
error between the actual image and the image predicted by
the model was below a predetermined threshold. In an
alternative embodiment, where there is only a
10 predetermined amount of time available for determining a
set of appearance parameters for an input image, this
iterative routine may be performed for a predetermined
period of time or for a predetermined number of
iterations.

CLAIMS:

1. A parametric model for modelling the shape of an object, the model comprising:

5 data defining a function which relates a set of input parameters to a set of locations which identify the relative positions of a plurality of predetermined points on the object;

10 characterised in that said data defines a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from one or more functions from a higher layer of the hierarchy and to generate therefrom at least some of said locations which identify the relative positions of said predetermined points.

15 20 2. A model according to claim 1, wherein said hierarchy comprises one or more intermediate layers of functions which are operable to receive parameters output from one or more functions from a higher layer of the hierarchy and to generate therefrom a set of output parameters for input to functions in a lower layer of the hierarchy.

25 3. A model according to claim 1 or 2, for modelling the two-dimensional shape of the object by identifying the

relative positions of said predetermined points in a predetermined plane.

4. A model according to claim 1 or 2, for modelling the
5 three-dimensional shape of the object by identifying the relative positions of the predetermined points in a three-dimensional space.

5. A model according to any preceding claim, wherein
10 one or more of said functions comprises a linear function which linearly relates the input parameters to the function to the output parameters of the function.

6. A model according to claim 5, wherein said one or
15 more linear functions are identified from a principal component analysis of training data derived from a set of training objects.

7. A model according to any preceding claim, wherein
20 one or more of said functions are non-linear.

8. A model according to claim 7, wherein at least one of said non-linear functions comprises a neural network.

25 9. A model according to any preceding claim, wherein the number of parameters input to at least one of said functions is smaller than the number of parameters output from the function.

10. A model according to any preceding claim, wherein the number of input parameters to at least one of said functions is greater than or equal to the number of parameters output by the function.

5

11. A model according to any preceding claim for modelling the shape and texture of the object, the model further comprising data defining a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from one or more functions from a higher layer of the hierarchy and to generate therefrom texture information for the object.

15

12. A model according to claim 11, wherein the texture hierarchy has the same structure as the shape hierarchy.

20

13. A model according to claim 11 or 12, wherein one or more of said functions are operable to relate an input set of shape and texture parameters to an output set of appearance parameters defining both shape and texture.

25

14. A model according to any preceding claim, wherein said object is a deformable object.

15. A model according to claim 14, wherein said

deformable object includes a human face.

16. A model according to claim 15, wherein said function in said top layer of the hierarchy models the shape of the entire face and wherein said hierarchy includes a 5 function which models the shape of the mouth.

17. A model according to claim 16, wherein said hierarchy further comprises a function for modelling the 10 shape of the eyes.

18. A model according to any preceding claim, wherein the or each function in the bottom layer of the hierarchy identifies the positions of a plurality of predetermined 15 points according to a predefined function of smaller number of control point positions.

19. A model according to claim 18, wherein the predefined function for each of the plurality of points 20 is a linear mapping of the control point positions and the control points are the three corners of a triangular facet.

20. A model according to claim 18, wherein the predefined function for each of the plurality of points 25 is a predefined non-linear mapping of a fixed number of control point positions.

21. A model according to claim 18, wherein the predefined function for each of the plurality of points is a predefined displacement from a single control point.

5 22. A method of determining a set of appearance parameters representative of the appearance of an object, the method comprising the steps of:

10 (i) storing a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object;

15 (ii) storing at least one function which relates a change in the input parameters to an error between actual appearance data for the object and appearance data determined from the set of input parameters and said parametric model;

20 (iii) initially estimating a current set of input parameters for the object;

(iv) determining appearance data for the object from the current set of input parameters and the stored parametric model;

(v) determining the error between actual appearance data of the object and the appearance data determined from the current set of input parameters;

25 (vi) determining a change in the input parameters using said at least one stored function and said determined error; and

(vii) updating the current set of input parameters

with the determined change in the input parameters.

23. A method according to claim 22, further comprising the step of repeating steps (iv) to (vii) until the error 5 determined in step (v) is less than a predetermined threshold.

24. A method according to claim 22, further comprising the step of repeating steps (iv) to (vii) for a predetermined amount of time or for a predetermined 10 number of repetitions.

25. A method according to claim 22, 23 or 24, wherein said second storing step stores a plurality of functions, 15 one associated with each function within the hierarchical model.

26. A method of tracking an object comprising the steps of:

20 (i) storing a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object;

25 (ii) storing at least one function which relates a change in the input parameters to an error between the actual appearance data for the object and the appearance data determined from the set of input parameters and said parametric model;

(iii) initially estimating a current set of input parameters for the object;

5 (iv) determining the appearance data for the object from the current set of input parameters and the stored parametric model;

(v) determining an error between the actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

10 (vi) determining a change in the input parameters using the at least one stored function and the determined error;

(vii) updating the current set of input parameters with said change in the input parameters;

15 (viii) repeating steps (iv) to (vii) in order to reduce the error determined in step (v); and

(ix) repeating steps (iii) to (viii) to track the object.

20 27. An apparatus for determining a set of appearance parameters representative of the appearance of an object, the apparatus comprising:

25 means for storing (i) a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object; and (ii) at least one function which relates a change in the input parameters to an error between actual appearance data for the object and

the appearance data for the object determined from the set of input parameters and said parametric model;

means for receiving an initial estimate of a current set of input parameters for the object;

5 means for updating the current set of input parameters comprising:

(i) means for determining appearance data for the object from the current set of input parameters and the stored parametric model;

10 (ii) means for determining the error between the actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

15 (iii) means for determining a change in the input parameters using said at least one stored function and said determined error; and

(iv) means for updating the current set of input parameters with the determined change in the input parameters.

20 28. An apparatus according to claim 27, wherein said updating means is operable to update iteratively the current set of input parameters until the error determining means determines an error which is less than a predetermined threshold.

25 29. An apparatus according to claim 27 or 28, wherein said storing means stores a plurality of functions, one

associated with each function within the hierarchical model.

30. An apparatus for tracking an object comprising:

5 means for storing (i) a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object; and (ii) at least one function which relates a change in the input parameters to an 10 error between actual appearance data for the object and the appearance data for the object determined from the set of input parameters and said parametric model;

15 means for receiving an initial estimate of a current set of input parameters for the object;

means for updating the current set of input parameters comprising:

(i) means for determining appearance data for the object from the current set of input parameters and the stored parametric model;

20 (ii) means for determining an error between actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

25 (iii) means for determining a change in the input parameters using the at least one stored function and the determined error; and

(iv) means for updating the current set of input parameters with said change in the input parameters;

wherein said updating means is operable to update iteratively the current set of input parameters in order to reduce the determined error, wherein said receiving means is operable to receive further estimates of the 5 current input parameters and wherein said update means is operable to update the received estimates of the current input parameters in order to track said object.

31. A storage medium storing the parametric model 10 according to any of claims 1 to 21 or storing processor implementable instructions for controlling a processor to implement the method of any one of claims 22 to 26.

32. Processor implementable instructions for controlling 15 a processor to implement the method of any one of claims 22 to 26.

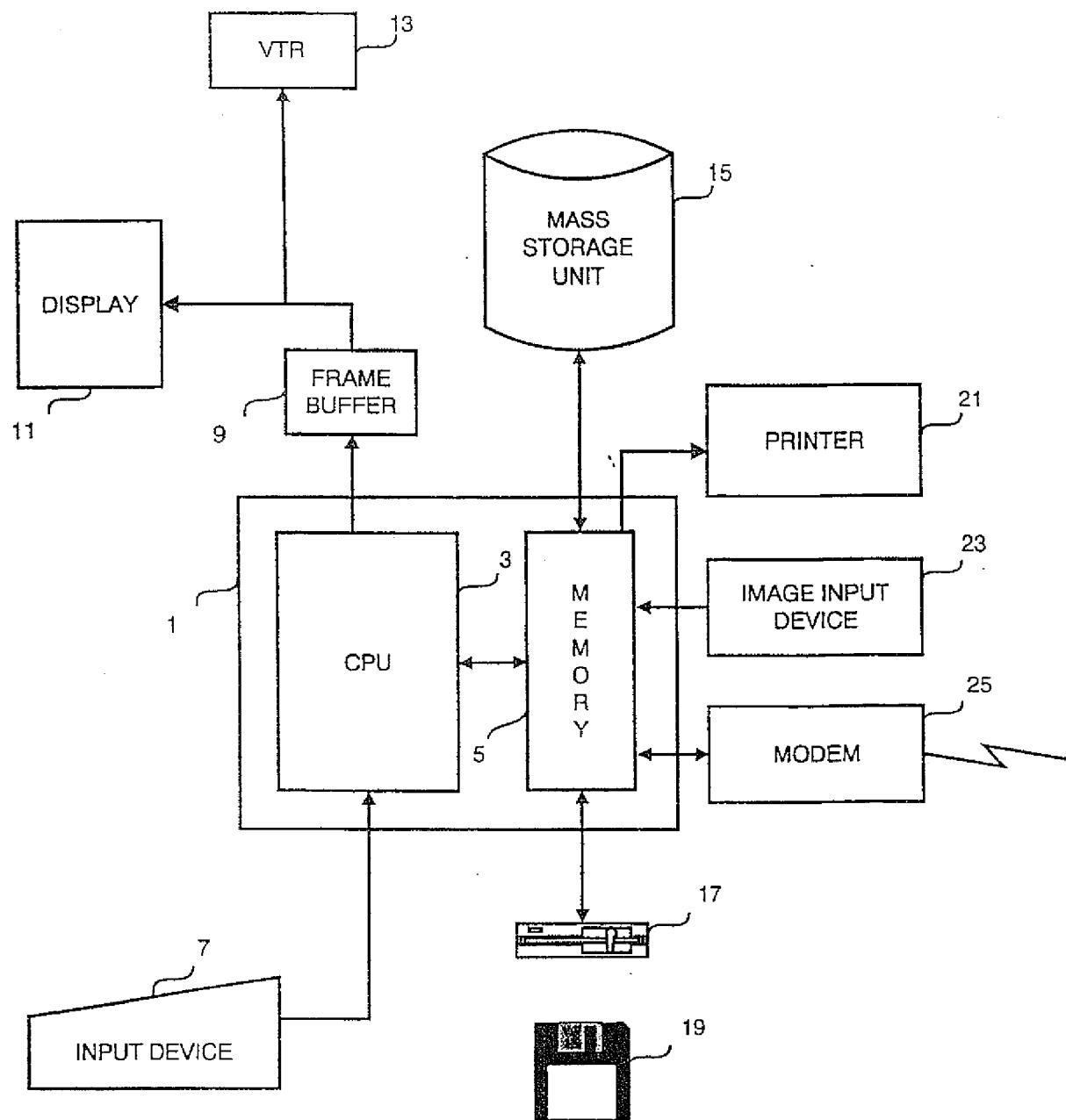


Fig. 1

2/14

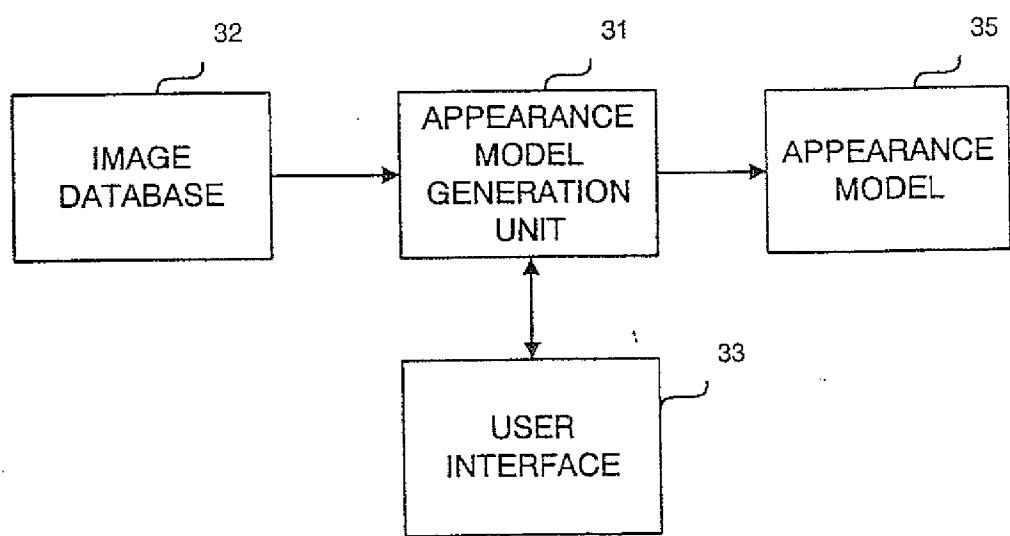


Fig. 2

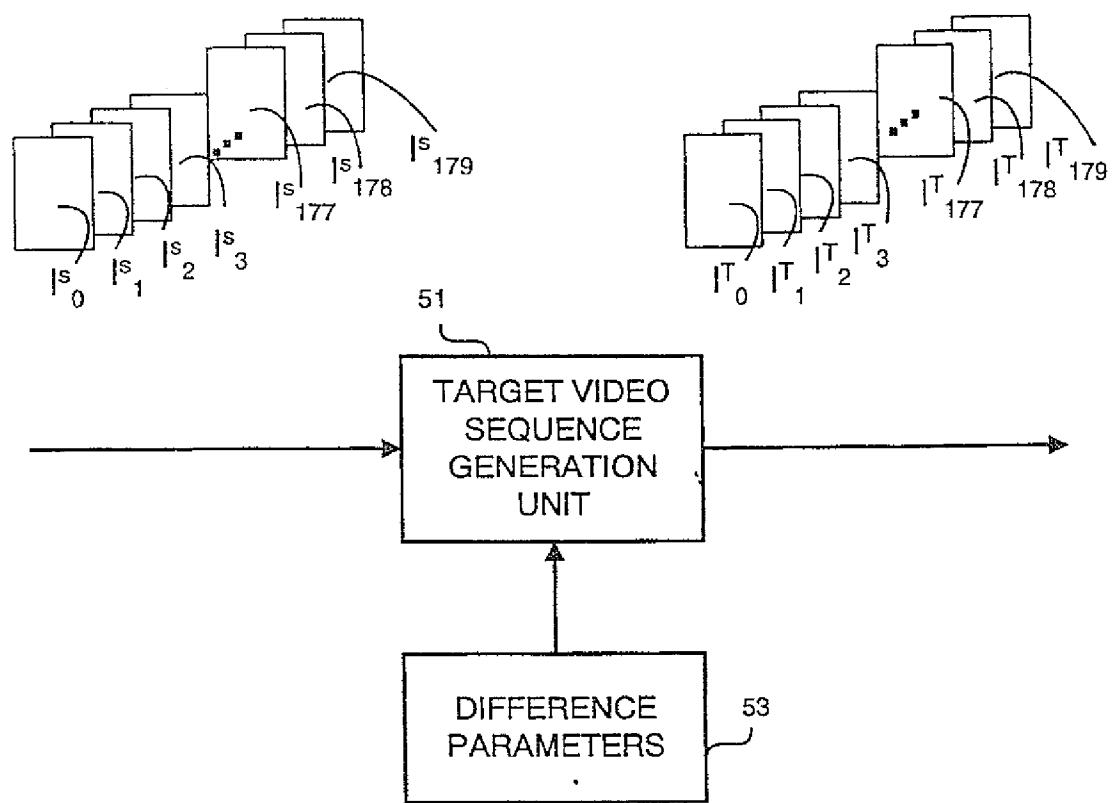


Fig. 3

4/14

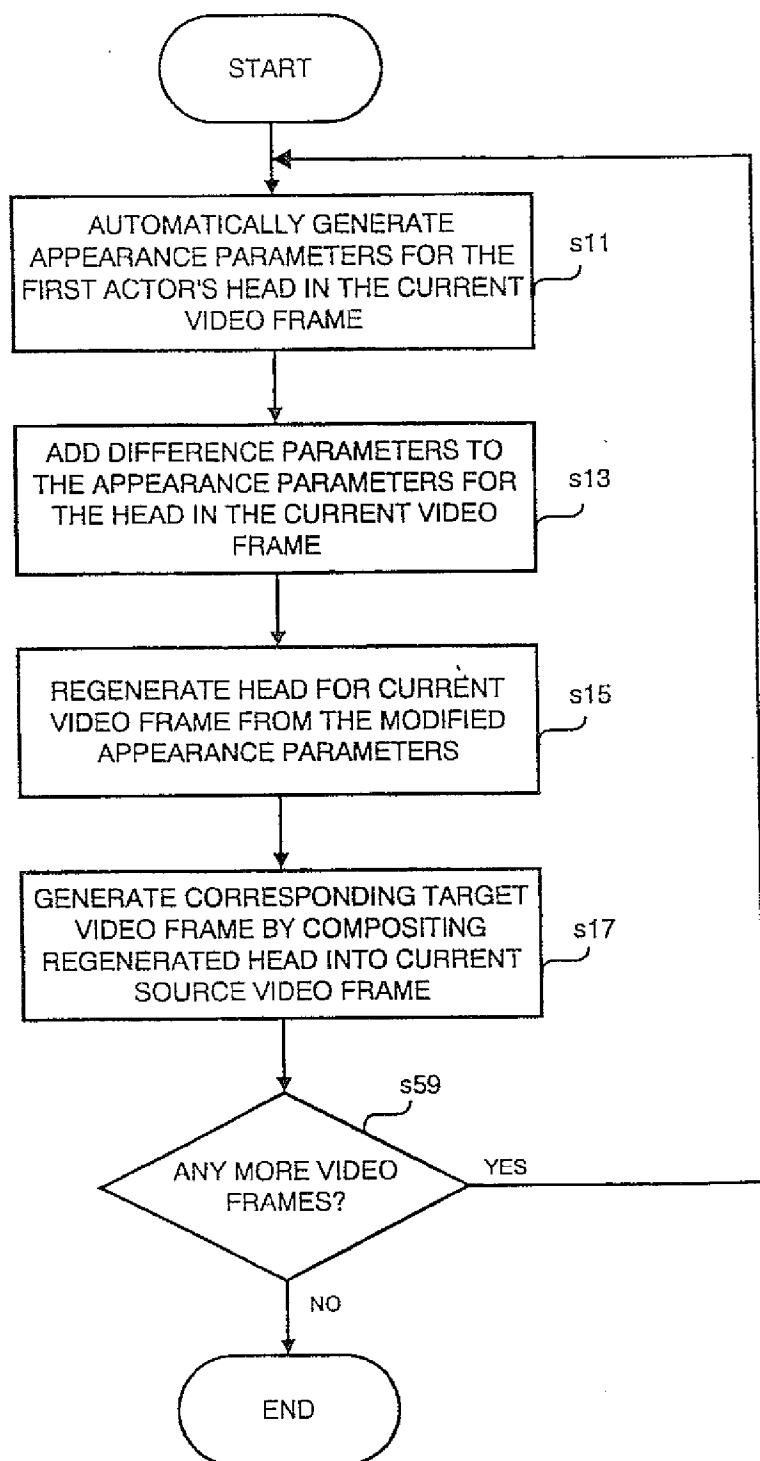


Fig. 4

5/14

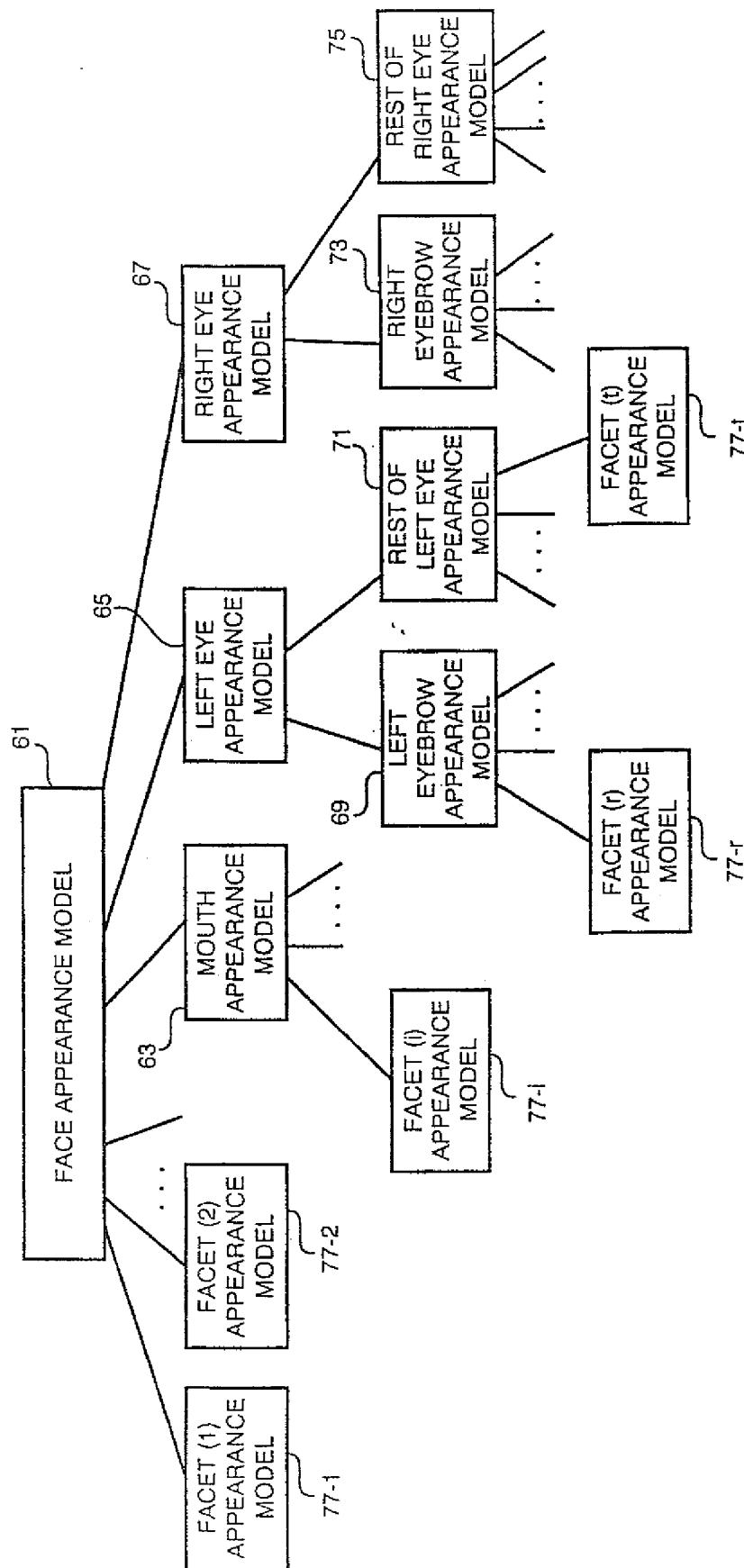


Fig. 5

6/14

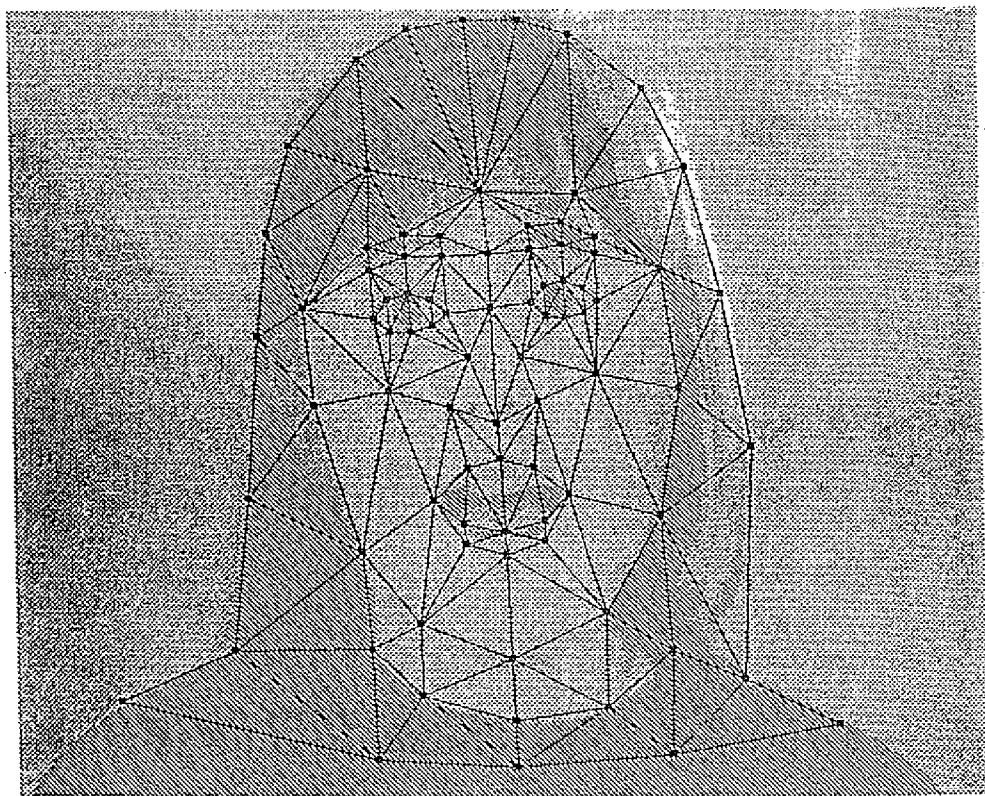


Fig. 6

7/14

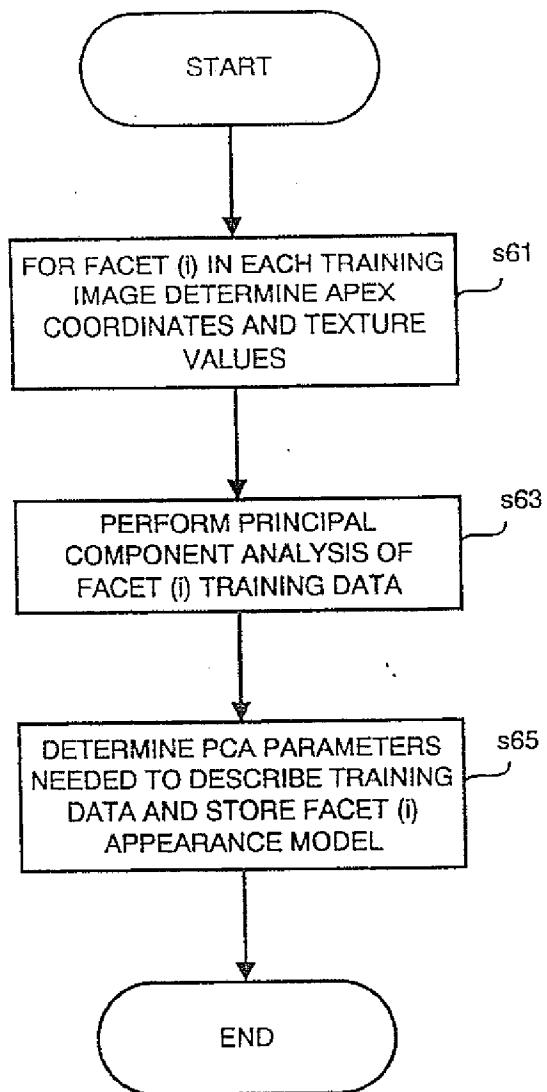


Fig. 7

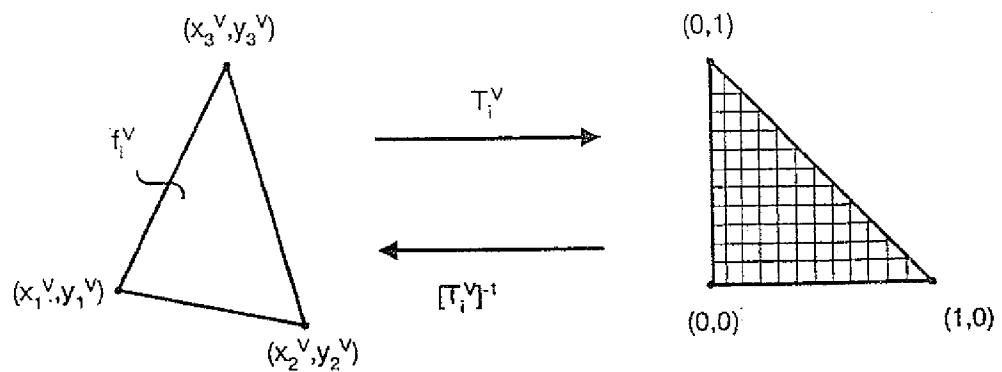


Fig. 8

9/14

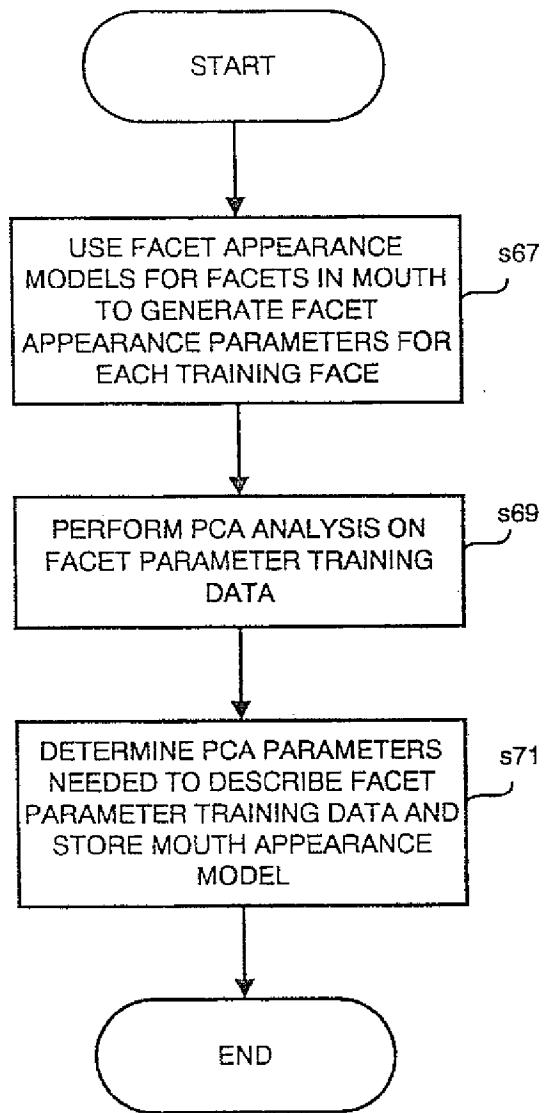
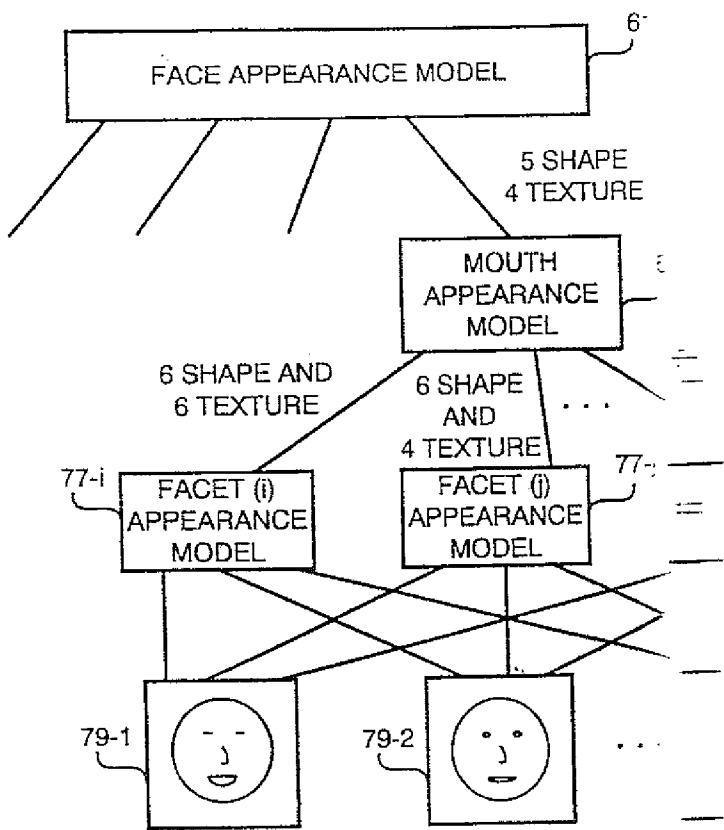


Fig. 9



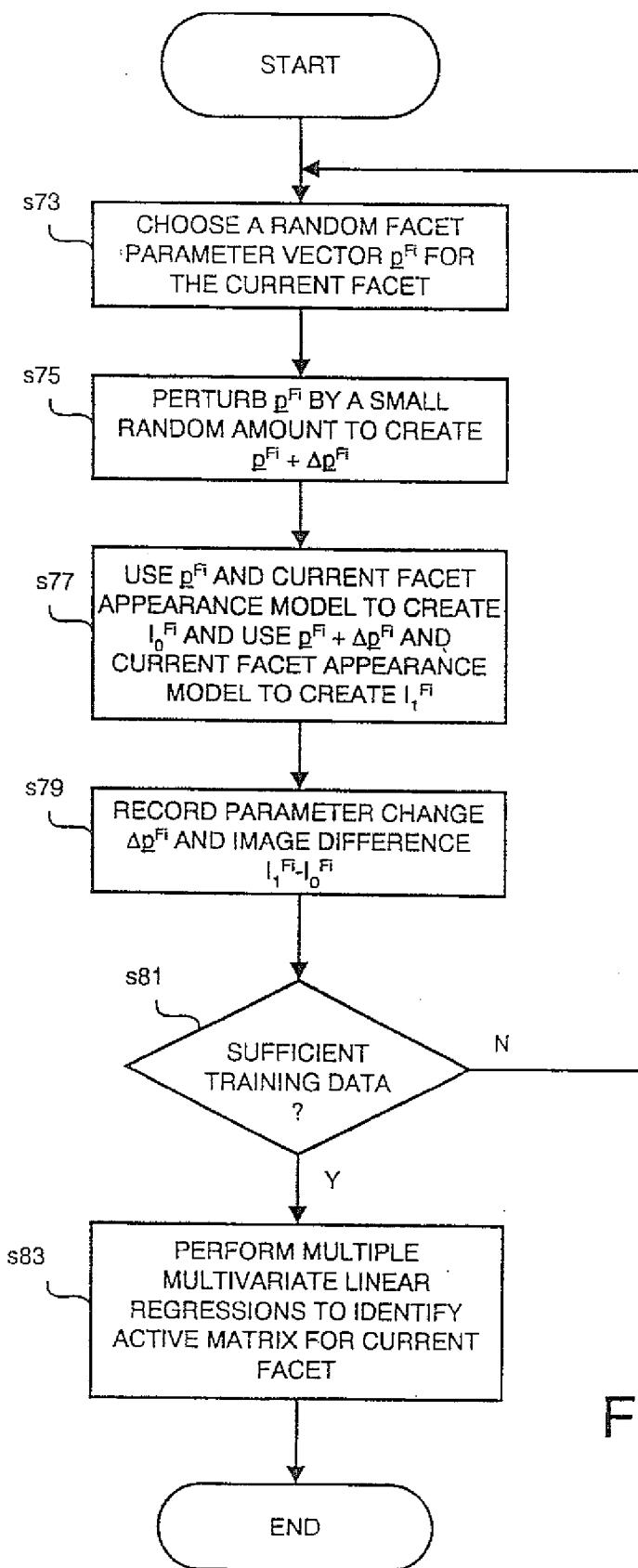


Fig. 11a

12/14

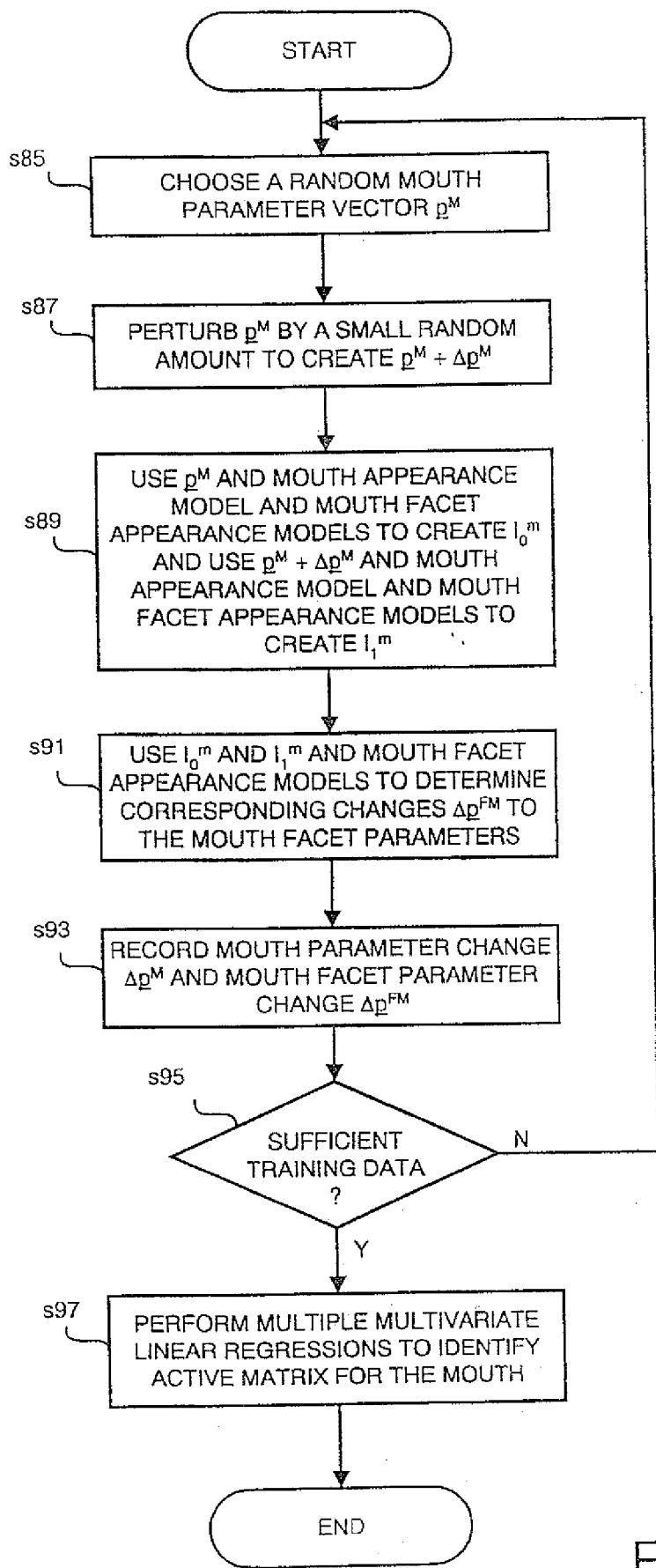


Fig. 11b

13/14

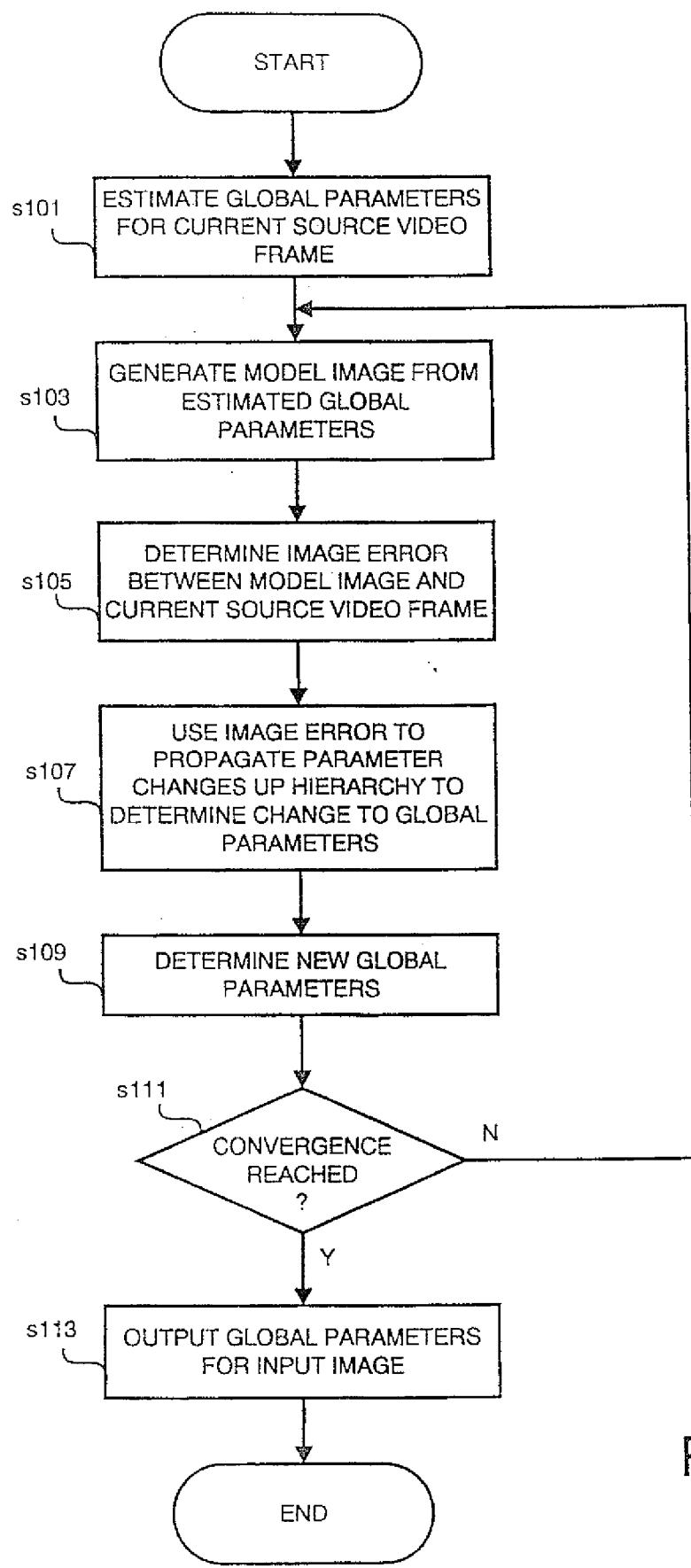


Fig. 12

14/14



Fig. 13a



Fig. 13b

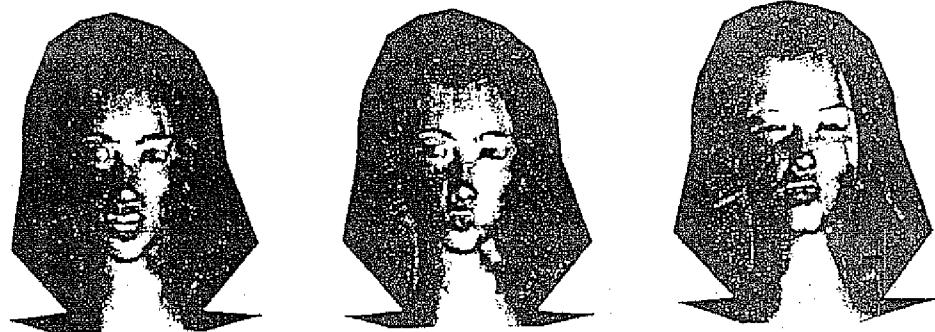


Fig. 13c



Fig. 13d



Fig. 13e

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 00/04411

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06T17/00 G06K9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06K G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC, IBM-TDB, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>EDWARDS G J ET AL: "ADVANCES IN ACTIVE APPEARANCE MODELS" KERKYRA, GREECE, SEPT. 20 - 27, 1999, LOS ALMITOS, CA: IEEE COMP. PRESS, US, vol. CONF. 7, 1999, pages 137-142, XP000980072 ISBN: 0-7695-0165-6 abstract page 137, left-hand column, paragraph 1 -right-hand column, paragraph 2</p> <p>-----</p> <p style="text-align: center;">-/-</p>	1-32

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

26 April 2001

Date of mailing of the international search report

04/05/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5518 Patentlaan 2
NL - 2280 HV Rijswijk

Authorized officer

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 00/04411

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	COOTES T F ET AL: "ACTIVE SHAPE MODELS-THEIR TRAINING AND APPLICATION" COMPUTER VISION AND IMAGE UNDERSTANDING, ACADEMIC PRESS, US, vol. 61, no. 1, January 1995 (1995-01), pages 38-59, XP000978654 ISSN: 1077-3142 abstract page 38, left-hand column, paragraph 1 -page 39, left-hand column, paragraph 1 -----	1-32

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/37222 A1(51) International Patent Classification⁷: G06T 17/00,
G06K 9/00W5 5EP (GB), WILLIAMS, Mark, Jonathan [GB/GB];
Anthropics Technology Limited, Ealing Studios, Ealing
Green, London W5 5EP (GB).

(21) International Application Number: PCT/GB00/04411

(74) Agents: BERESFORD, Keith, Denis, Lewis et al.; Beresford & Co, 2-5 Warwick Court, High Holborn, London
WC1R 5DJ (GB).

(22) International Filing Date:

20 November 2000 (20.11.2000)

(25) Filing Language:

English

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(26) Publication Language:

English

(30) Priority Data:

9927314.6 18 November 1999 (18.11.1999) GB

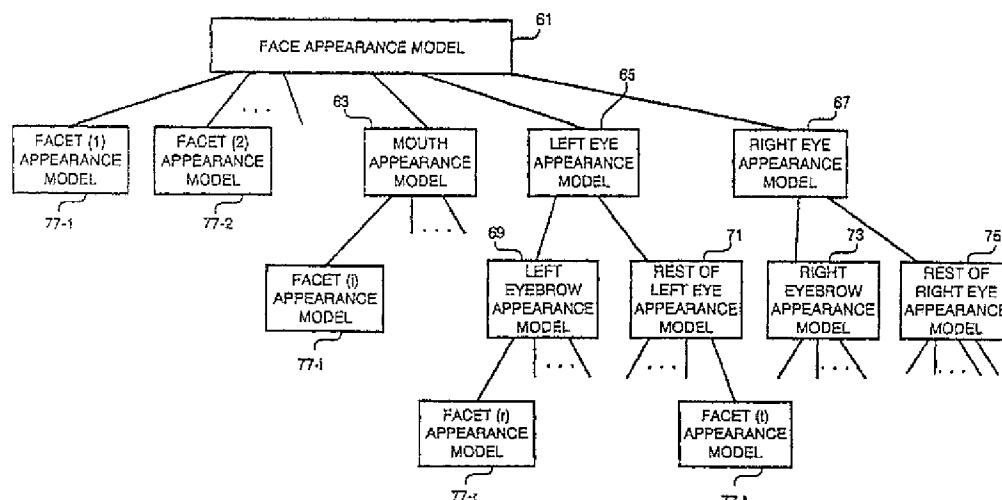
(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).(71) Applicant (for all designated States except US): AN-
THROPICS TECHNOLOGY LIMITED [GB/GB];
Ealing Studios, Ealing Green, London W5 5EP (GB).

Published:

— with international search report

[Continued on next page]

(54) Title: IMAGE PROCESSING SYSTEM



(57) Abstract: A hierarchical parametric model is provided for modelling the appearance of objects, such as human faces. The hierarchical model can model both the shape and the texture of the object. The hierarchical model includes models for components of the object being modelled such that output parameters from one model are applied as input parameters to models which are lower in the hierarchy. This hierarchical model can be used, for example, for face tracking, for video compression, for 2D and 3D character

WO 01/37222 A1



(48) Date of publication of this corrected version:

9 August 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(15) Information about Correction:

see PCT Gazette No. 32/2001 of 9 August 2001, Section II

IMAGE PROCESSING SYSTEM

5 The present invention relates to the parametric modelling of the appearance of objects. The resulting model can be used, for example, to track the object, such as a human face, in a video sequence.

10 The use of parametric models for image interpretation and synthesis has become increasing popular. Cootes et al have shown in their paper entitled "Active Shape Models - Their Training and Application", Computer Vision and Image Understanding, Volume 61, No. 1, January, pages 38-59, 1995, how such parametric models can be used to model the variability of the shape and texture of human faces.

15 They have mainly used these models for face recognition and tracking within video sequences, although they have also demonstrated that their model can be used to model the variability of other deformable objects, such as MRI scans of knee joints. The use of these models provides a basis for a broad range of applications since they explain the appearance of a given image in terms of a compact set of model parameters which can be used for higher levels of interpretation of the image. For example, when analysing face images, they can be used to 20 characterise the identity, pose or expression of a face.

25

Using such models for image interpretation requires, however, a method of fitting them to new image data.

This involves identifying the model parameters that generate an image which best fits (according to some measure) the new input image. Typically this problem is one of minimising the sum of squares of pixel errors between the generated image and the input image. In their paper entitled "Estimating Coloured 3D Face Models from Single Images: An Example-Based Approach" Vetter and Blanz have proposed a stochastic gradient descent optimisation technique to identify the optimum model parameters for the new image. Although this technique can give very accurate results finding the locally optimal solution, they generally get stuck in local minima since the error surface for the problem of fitting an appearance model to an image is particularly rough containing many local minima. Therefore, this minimisation technique often fails to converge on the global minimum. An additional drawback of this technique is that it is very slow requiring several minutes to achieve convergence.

20

A faster, more robust technique known as the active appearance model was proposed by Edwards et al in the paper entitled "Interpreting Face Images using Active Appearance Models", published in the Third International Conference on Automatic Face and Gesture Recognition 1998, pages 300-305, Japan, April 1998. This technique uses a prior training stage in which the relationship between model parameter displacements and the resulting

change in image error is learnt. Although the method is much faster than direct optimisation techniques, it also requires fairly accurate initial model parameters if the search is to converge. Additionally, this technique does 5 not guarantee that the optimum parameters will be found.

The appearance model proposed by Cootes et al includes a single appearance model matrix which linearly relates a set of parameters to corresponding image data. Blanz et 10 al segmented the face into a number of completely independent appearance models, each of which is used to render a separate region of the face. The results are then merged using a general interpretation technique.

15 The present invention aims to provide an alternative way of modelling the appearance of objects which will allow subsequent image interpretation through appropriate processing of parameters generated for the image.

20 According to one aspect, the present invention provides a hierarchical parametric model for modelling the shape of an object, the model comprising data defining a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set 25 of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from one or more functions from a higher layer of the

hierarchy and to generate therefrom the relative positions of a plurality of predetermined points on the object. Such a hierarchical parametric model has the advantage that small changes in some parts of the object 5 can still be modelled by the parameters, even though they are significantly smaller than variations in other less important parts of the object. This model can be used for face tracking, video compression, 2D and 3D character generation, face recognition for security purposes, image 10 editing etc.

According to another aspect, the present invention provides an apparatus and method of determining a set of appearance parameters representative of the appearance of 15 an object, the method comprising the steps of storing a hierarchical parametric model such as the one discussed above and at least one function which relates a change in input parameters to an error between actual appearance data for the object and appearance data determined from 20 the set of input parameters and the parametric model; initially receiving a current set of input parameters for the object; determining appearance data for the object from the current set of input parameters and the stored parametric model; determining the error between the 25 actual appearance data of the object and the appearance data determined from the current set of input parameters; determining a change in the input parameters using the at least one stored function and said determined error; and

updating the current set of input parameters with the determined change in the input parameters.

5 An exemplary embodiment of the present invention will now be described with reference to the accompanying drawings in which:

10 Figure 1 is a schematic block diagram illustrating a general arrangement of a computer system which can be programmed to implement the present invention;

15 Figure 2 is a block diagram of an appearance model generation unit which receives some of the image frames of a source video sequence together with a target image frame and generates therefrom an appearance model;

20 Figure 3 is a block diagram of a target video sequence generation unit which generates a target video sequence from a source video sequence using a set of stored difference parameters;

25 Figure 4 is a flow chart illustrating the processing steps which the target video sequence generation unit shown in Figure 3 performs to generate the target video sequence;

Figure 5 schematically illustrates the form of a hierarchical appearance model generated in one embodiment

of the invention;

5 Figure 6 shows a head with a mesh of triangular facets placed over the head and whose positions are defined by the position of landmark points at the corners of the facets;

10 Figure 7 is a flow chart illustrating the processing steps required to generate a facet appearance model from the training images;

15 Figure 8 schematically illustrates the way in which a transformation is defined between a facet in a training image and a predefined shape of facet which allows texture information to be extracted from the facet;

20 Figure 9 is a flow chart illustrating the main processing steps involved in determining an appearance model for the mouth using the appearance models for the facets which appear in the mouth and using the training images;

25 Figure 10 schematically illustrates the way in which training images are used to determine some of the appearance models which form the hierarchical appearance model illustrated in Figure 5;

Figure 11a is a flow chart illustrating the processing steps performed during a training routine to identify an

Active matrix associated with a current facet;

Figure 11b is a flow chart illustrating the processing steps performed during a training routine to identify an Active matrix associated with the mouth;

10

Figure 12 is a flow chart illustrating the processing steps involved in determining a set of parameters which define the appearance of a face within a input image;

15

Figure 13a shows three frames of an example source video sequence which is applied to the target video sequence generation unit shown in Figure 4;

20

Figure 13b shows an example target image used to generate a set of difference parameters used by the target video sequence generation unit shown in Figure 4;

25

Figure 13c shows a corresponding three frames from a target video sequence generated by the target video sequence generation unit shown in Figure 4 from the three frames of the source video sequence shown in Figure 13a using the difference parameters generated using the target image shown in Figure 13b;

Figure 13d shows a second example of a target image used to generate a set of difference parameters for use by the target video sequence generation unit shown in Figure 4;

and

5 Figure 13e shows the corresponding three frames from the target video sequence generated by the target video sequence generation unit shown in Figure 4 when the three frames of the source video sequence shown in Figure 13a are input to the target video sequence generation unit together with the difference parameters calculated using the target image shown in Figure 13d.

10

Figure 1 is an image processing apparatus according to an embodiment of the present invention. The apparatus comprises a computer 1 having a central processing unit (CPU) 3 connected to a memory 5 which is operable to store a program defining the sequence of operations of the CPU 3 and to store object and image data used in calculations by the CPU 3. Coupled to an input port of the CPU 3 there is an input device 7, which in this embodiment comprises a keyboard and a computer mouse. Instead of, or in addition to the computer mouse, another position sensitive input device (pointing device) such as a digitiser with associated stylus may be used.

20
25

A frame buffer 9 is also provided and is coupled to the CPU 3 and comprises a memory unit (not shown) arranged to store image data relating to at least one image, for example by providing one (or several) memory location(s) per pixel of the image. The value stored in the frame

buffer for each pixel defines the colour or intensity of that pixel in the image. In this embodiment, the images are represented by 2-D arrays of pixels, and are conveniently described in terms of Cartesian coordinates, 5 so that the position of a given pixel can be described by a pair of x-y coordinates. This representation is convenient since the image is displayed on a raster scan display 11. Therefore, the x-coordinate maps to the distance along the line of the display and the y-coordinate maps to the number of the line. The frame 10 buffer 9 has sufficient memory capacity to store at least one image. For example, for an image having a resolution of 1000 x 1000 pixels, the frame buffer 9 includes 10^6 pixel locations, each addressable directly or indirectly 15 in terms of a pixel coordinate x,y.

In this embodiment, a video tape recorder (VTR) 13 is also coupled to the frame buffer 9, for recording the image or sequence of images displayed on the display 11. 20 A mass storage device 15, such as a hard disc drive, having a high data storage capacity is also provided and coupled to the memory 5. Also coupled to the memory 5 is a floppy disc drive 17 which is operable to accept removable data storage media, such as a floppy disc 19 25 and to transfer data stored thereon to the memory 5. The memory 5 is also coupled to a printer 21 so that generated images can be output in paper form, an image input device 23 such as a scanner or video camera and a

modem 25 so that input images and output images can be received from and transmitted to remote computer terminals via a data network, such as the Internet. The CPU 3, memory 5, frame buffer 9, display unit 11 and mass storage device 13 may be commercially available as a complete system, for example as an IBM compatible personal computer (PC) or a workstation such as the Sparc station available from Sun Microsystems.

10 A number of embodiments of the invention can be supplied commercially in the form of programs stored on a floppy disc 19 or on other mediums, or as signals transmitted over a data link, such as the Internet, so that the receiving hardware becomes reconfigured into an apparatus 15 embodying the present invention.

In this embodiment, the computer 1 is programmed to receive a source video sequence input by the image input device 23 and to generate a target video sequence from the source video sequence using a target image. In this embodiment, the source video sequence is a video clip of 20 an actor acting out a scene, the target image is an image of a second actor and the resulting target video sequence is a video sequence showing the second actor acting out 25 the scene. The way in which this is achieved will now be briefly described with reference to Figures 2 to 4.

In this embodiment, in order to generate the target video

sequence from the source video sequence, a hierarchical parametric appearance model which models the variability of shape and texture of the head images is used. This appearance model makes use of the fact that some prior knowledge is available about the contents of head images in order to facilitate their modelling. For example, it can be assumed that two frontal images of a human face will each include eyes, a nose and a mouth. In this embodiment, as shown in Figure 2, the hierarchical parametric appearance model 35 is generated by an appearance model generation unit 31 from training images which are stored in an image database 32. In this embodiment, all the training images are colour images having 500 x 500 pixels, with each pixel having a red, green and a blue pixel value. The resulting appearance model 35 is a parameterisation of the appearance of the class of head images defined by the heads in the training images, so that a relatively small number of parameters (for example 50) can describe the detailed (pixel level) appearance of a head image from the class. In particular, the hierarchical appearance model 35 defines a function (F) such that:

$$I = F(p) \quad (1)$$

where p is the set of appearance parameters (written in vector notation) which generates, through the hierarchical appearance model (F), the face image I . The

structure of the hierarchical appearance model used in this embodiment will be described later.

Once the hierarchical appearance model 35 has been 5 determined, a target video sequence can be generated from a source video sequence. As shown in Figure 3, the source video sequence is input to a target video sequence generation unit 51 which processes the source video sequence using a set of difference parameters 53 to generate and to output the target video sequence. The 10 difference parameters 53 are determined by subtracting the appearance parameters which are generated for the first actor's head in one of the source video frames, from the appearance parameters which are generated for 15 the second actor's head in the target image. The way in which these appearance parameters are determined for these images will be described later. In order that these difference parameters only represent differences in the general shape and colour texture of the two actors' 20 heads, the pose and facial expression of the first actor's head in the source video frame used should match, as closely as possible, the pose and facial expression of the second actor's head in the target image.

25 The processing steps required to generate the target video sequence from the source video sequence will now be described in more detail with reference to Figure 4. As shown, in step s1, the appearance parameters (p_s^i) for

the first actor's head in the current video frame (I_s^i) are automatically calculated. The way that this is achieved will be described later. Then, in step s3, the difference parameters (p_{dif}) are added to the appearance parameters for the first actor's head in the current video frame to generate:

$$p_{mod}^i = p_s^i + p_{dif} \quad (2)$$

10 The resulting appearance parameters (p_{mod}^i) are then used, in step s5, to regenerate the head for the current target video frame. In particular, the modified appearance parameters are inserted into equation (1) above to regenerate a modified head image which is then 15 composited, in step s7, into the source video frame to generate the corresponding target video frame. A check is then made, in step s9, to determine whether or not there are any more source video frames. If there are, then the processing returns to step s1 where the 20 procedure described above is repeated for the next source video frame. If there are no more source video frames, then the processing ends.

Figure 13 illustrates the results of this animation technique (although showing black and white images and 25 not colour). In particular, Figure 13a shows three frames of the source video sequence, Figure 13b shows the target image (which in this embodiment is computer

generated) and Figure 13c shows the corresponding three frames of the target video sequence obtained in the manner described above. As can be seen, an animated sequence of the computer generated character has been 5 generated from a video clip of a real person and a single image of the computer generated character.

HIERARCHICAL APPEARANCE MODEL

In the systems described by Cootes et al and Blanz et al, 10 the parametric model is created by placing a number of landmark points on a training image and then identifying the same landmark points on the other training images in order to identify how the location of and the pixel values around the landmark points vary within the 15 training images. A principal component analysis is then performed on the matrix which consists of vectors of the landmark points. This PCA yields a set of Eigenvectors which describe the directions of greatest variation along which the landmark points change. Their appearance model 20 includes the linear combination of the Eigenvectors plus parameters for translation, rotation and scaling. This single appearance model relates a compact set of appearance parameters to pixel values.

25 In this embodiment, rather than having a single appearance model for the object, a hierarchical appearance model comprising several appearance models which model variations in components of the object is

used. For example, in the case of human faces, the hierarchical appearance model may include an appearance model for the mouth, one for the left eye, one for the right eye and one for the nose. Since it may be possible to model various components of the object, the particular hierarchical structure which will be used for a particular object and application must first of all be defined by the system designer.

Figure 5 schematically illustrates the structure of the hierarchical appearance model used in this embodiment. As shown, at the top of the hierarchy there is a general face appearance model 61. Beneath the face appearance model there is a mouth appearance model 63, a left eye appearance model 65, a right eye appearance model 67, a left eyebrow appearance model 69, a rest of left eye appearance model 71, a right eyebrow appearance model 73, a rest of right eye appearance model 75 and, in this embodiment, a facet appearance model for each facet defined in the training images. Figure 6 shows the head of a training image in which the set of landmark points has been placed at the appropriate points on the head. As shown, in this embodiment, there are one hundred and forty-eight triangular areas or facets defined by the positions of the landmark points. Therefore, in this embodiment, there are one hundred and forty-eight facet appearance models 77.

The face appearance model 61 operates to relate a small number of "global" appearance parameters to a further set of appearance parameters, some of which are input to facet appearance models 77, some of which are input to the mouth appearance model 63, some of which are input to the left eye appearance model 65 and the rest of which are input to the right eye appearance model 67. The facet appearance models 77 operate to relate the input parameters received from the appearance model which is above it in the hierarchy into corresponding pixel values for that facet. The mouth appearance model 63 is operable to relate the parameters it receives from the face appearance model 61 into a further set of appearance parameters, respective ones of which are output to the respective facet appearance models 77 for the facets which are associated with the mouth. Similarly, the left and right eye appearance models 65 and 67 operate to relate the parameters it receives from the face appearance model 61 into a further set of appearance parameters, some of which are input to the appropriate eyebrow appearance model and the rest of which are input to the appropriate rest of eye appearance model. These appearance models in turn convert these parameters into parameters for input to the facet appearance models associated with the facets which appear in the left and right eyes respectively. In this way, a small compact set of "global" appearance parameters input to the face appearance model 61 can filter through the hierarchical

structure illustrated in Figure 5 to generate a set of pixel values for all the facets in a head which can then be used to regenerate the image of the head.

5 The way in which the individual appearance models of this hierarchical appearance model are generated in this embodiment will now be described with reference to Figures 6 to 10.

10 In this embodiment, each of the training images stored in the image database 32 is labelled with eighty six landmark points. In this embodiment, this is performed manually by the user via the user interface 33. In particular, each training image is displayed on the display 11 and the user places the landmark points over the head in the training image. These points delineate the main features in the head, such as the position of the hairline, neck, eyes, nose, ears and mouth. In order to compare training faces, each landmark point is 15 associated with the same point on each face. In this 20 embodiment, the following landmark points are used:

Landmark Point	Associated Position	Landmark Point	Associated Position
LP ₁	Left corner of left eye	LP ₄₄	Eye, bottom
LP ₂	Right corner of right eye	LP ₄₅	Eye, top
LP ₃	Chin, bottom	LP ₄₆	Eye, bottom
LP ₄	Right corner of left eye	LP ₄₇	Eyebrow, lower

Landmark Point	Associated Position	Landmark Point	Associated Position
LP ₅	Left corner of right eye	LP ₄₈	Eyebrow, upper
LP ₆	Mouth, left	LP ₄₉	Cheek, left
LP ₇	Mouth, right	LP ₅₀	Cheek, right
LP ₈	Nose, bottom	LP ₅₁	Eyebrow, lower
5 LP ₉	Nose, between eyes	LP ₅₂	Eyebrow, upper
LP ₁₀	Upper lip, top	LP ₅₃	Eyebrow, lower
LP ₁₁	Lower lip, bottom	LP ₅₄	Eyebrow, upper
LP ₁₂	Neck, left, top	LP ₅₅	Eyebrow, lower
LP ₁₃	Neck, right, top	LP ₅₆	Eyebrow, upper
10 LP ₁₄	Face edge left, level with nose	LP ₅₇	Eyebrow, lower
LP ₁₅	Face edge	LP ₅₈	Eyebrow, upper
LP ₁₆	Face edge right, level with nose	LP ₅₉	Eyebrow, lower
LP ₁₇	Face edge	LP ₆₀	Eyebrow, upper
LP ₁₈	Top of head	LP ₆₁	Eyebrow, lower
15 LP ₁₉	Hair edge	LP ₆₂	Lower lip, top
LP ₂₀	Hair edge	LP ₆₃	Centre forehead
LP ₂₁	Hair edge	LP ₆₄	Upper lip, top left
LP ₂₂	Hair edge	LP ₆₅	Upper lip, top right
LP ₂₃	Hair edge	LP ₆₆	Lower lip, bottom right
20 LP ₂₄	Hair edge	LP ₆₇	Lower lip, bottom left
LP ₂₅	Hair edge	LP ₆₈	Eye, top left
LP ₂₆	Hair edge	LP ₆₉	Eye, top right
LP ₂₇	Hair edge	LP ₇₀	Eye, bottom right
LP ₂₈	Hair edge	LP ₇₁	Eye, bottom left
25 LP ₂₉	Bottom, far left	LP ₇₂	Eye, top left
LP ₃₀	Bottom, far right	LP ₇₃	Eye, top right
LP ₃₁	Shoulder	LP ₇₄	Eye, bottom right

Landmark Point	Associated Position	Landmark Point	Associated Position
LP ₃₂	Shoulder	LP ₇₅	Eye, bottom left
LP ₃₃	Bottom, left	LP ₇₆	Lower lip, top left
LP ₃₄	Bottom, middle	LP ₇₇	Lower lip, top right
LP ₃₅	Bottom, right	LP ₇₈	Chin, left
LP ₃₆	Left forehead	LP ₇₉	Chin, right
LP ₃₇	Right forehead	LP ₈₀	Neck, left
LP ₃₈	Centre, between eyebrows	LP ₈₁	Neckline, left
LP ₃₉	Nose, left	LP ₈₂	Neckline
LP ₄₀	Nose, right	LP ₈₃	Neckline, right
LP ₄₁	Nose edge, left	LP ₈₄	Neck, right
LP ₄₂	Nose edge, right	LP ₈₅	Hair edge
LP ₄₃	Eye, top	LP ₈₆	Hair edge

The result of the manual placement of the landmark points
15 is a table of landmark points for each training image,
which identifies the (x, y) coordinate of each landmark
point within the image. As shown in Figure 6, these
landmark points are also used to define the location of
predetermined triangular facets or areas within the
20 training image.

FACT APPEARANCE MODEL

Figure 7 shows a flow chart illustrating the main
processing steps involved in this embodiment in
25 determining a facet appearance model for facet (i). As
shown, in step s61, the system determines, for each
training image, the apex coordinates of facet (i) and

texture values from within facet (i). In order to sample texture from within the facet at corresponding points within each training facet, a transformation which transforms the facet onto a reference facet is determined. Figure 8 illustrates this transformation. In particular, Figure 8 shows facet f_i^v taken from the v -th training image, which is defined by the landmark points (x_1^v, y_1^v) , (x_2^v, y_2^v) and (x_3^v, y_3^v) . The transformation (T_i^v) which transforms those coordinates onto coordinates $(0,0)$, $(1,0)$ and $(0,1)$ is determined. In this embodiment, the texture information extracted from each training facet is defined by the regular array of pixels shown in the reference facet. In order to determine the corresponding red, green and blue pixel values in the training image, the inverse transformation $([T_i^v]^{-1})$ is used to transform the pixel locations in the reference facet, into corresponding locations in the training facet, from which the RGB pixel values are determined. In this embodiment, this transformation may not result in an exact correspondence with a single image pixel location since the pixel resolution in the actual facet may be different to the resolution in the reference facet. In this embodiment, the texture information (RGB pixel values) which is determined is obtained by interpolating between the surrounding image RGB pixel values. In this embodiment, there are fifty pixels in the regular array of pixels in the reference facet. Therefore, fifty RGB pixel values are extracted for each

training facet. The texture information for facet (i) from the v-th training image can then be represented by a vector (t^{iv}) of the form:

5

$$\underline{t}^{iv} = [t_1^{iv}, t_2^{iv}, t_3^{iv} \dots t_{50}^{iv}]^T$$

where t_i^{iv} is the RGB texture information for the first reference pixel extracted from facet (i) in the v-th training image etc.

10

15

20

25

In this embodiment, the facet appearance models 77 treat shape and texture separately. Therefore, in step s63, the system performs a principal component analysis (PCA) on the set of texture training vectors generated in step s61. For a more detailed discussion of principal component analysis, the reader is referred to the book by W. J. Krzanowski entitled "Principles of Multivariate Analysis - A User's Perspective" 1998, Oxford Statistical Science Series. As those skilled in the art will appreciate, this principal component analysis determines all possible modes of variation within the training texture vectors. However, since each of the facets is associated with a similar point on the face, most of the variation within the data can be explained by a few modes of variation. The result of the principal component analysis is a facet texture appearance model (defined by matrix F_1) which relates a vector of facet texture parameters to a vector of texture pixel values, by:

$$p_v^{fit} = F_i (\underline{t}^{iv} - \underline{\bar{t}}^i) \quad (3)$$

where \underline{t}^{iv} is the RGB texture vector defined above, $\underline{\bar{t}}^i$ is the mean RGB texture vector for facet (i), F_i is a matrix which defines the facet texture appearance model for facet (i) and p_v^{fit} is a vector of the facet texture parameters which describes the RGB texture vector \underline{t}^{iv} . The matrix F_i describes the main modes of variation of the texture within the training facets; and the vector of facet texture parameters (p_v^{fit}) for a given input facet has a parameter associated with each mode of variation whose value relates the texture of the input facet to the corresponding mode of variation.

As those skilled in the art will appreciate, for facets which describe fairly constant parts of the face, such as the chin or cheeks, very few parameters will be needed to model the variability within the training images. However, facets which are associated with areas of the face where there is a large amount of variability (such as facets which form part of the eye), will require a larger number of facet texture parameters to describe the variability within the training images. Therefore, in step 565, the system determines how many texture parameters are needed for the current facet and stores the appropriate facet appearance model matrix.

In addition to being able to determine a set of texture

parameters p^{fit}_v for a given texture vector \underline{t}^{iv} , equation (3) can be solved with respect to the texture vector \underline{t}^{iv} to give:

5.
$$\underline{t}^{iv} = \underline{t}^i - F_i^T D_v^{Fit} \quad (4)$$

since $F_i F_i^T$ equals the identity matrix. Therefore, by modifying the set of texture parameters (p^{fit}) within suitable limits, new textures for facet (i) can be generated which are similar to those in the training set.

10 Once the above procedure has been performed for each of the one hundred and forty-eight facets in the training images, a facet texture appearance model will have been generated for each of those facets. In this embodiment, 15 the facet appearance model does not compress the parameters defining the shape of the facets, since only six parameters are needed to define the shape of each facet - two parameters for each (x,y) coordinate of the 20 facet's apexes.

MOUTH APPEARANCE MODEL

Figure 9 shows a flow chart illustrating the main processing steps required in order to generate the mouth appearance model 63. As shown, in step s67, the system 25 uses the facet appearance models for the facets which form part of the mouth to generate shape and texture parameters from those facets for each training image.

Therefore, referring to Figure 10, the mouth appearance model 63 will receive texture and shape parameters from the facet appearance model for facet (i), facet (j) and facet (n) for the corresponding facets in each of the 5 training images 79. As illustrated in Figure 10, the appearance model for facet (i) is operable to generate, for each training image, six shape parameters (corresponding to the three (x,y) coordinates of the apexes of facet (i)) and six texture parameters. 10 Similarly, the appearance model for facet (j) is operable to generate, for each training image, six shape parameters and four texture parameters and the appearance model for facet (n) is operable to generate, for each training image, six shape parameters and three texture 15 parameters.

The processing then proceeds to step s69 where the system performs a principal component analysis on the shape and texture parameters generated for the training images by 20 the facet appearance models associated with the mouth. In this embodiment, the mouth appearance model 63 treats the shape and texture separately. In particular, for each training image, the system concatenates the six shape parameters for the facets associated with the mouth 25 to form the following shape vector:

$$p^{FMS} = [x_1^{fi}, y_1^{fi}, x_2^{fi}, y_2^{fi}, x_3^{fi}, y_3^{fi} : x_1^{fj}, y_1^{fj}, x_2^{fj}, y_2^{fj} \dots]^T$$

and concatenates the facet texture parameters output by the facet appearance models associated with the mouth to form the following texture vector:

5 $\mathbf{p}^{\text{Mt}} = [p_1^{\text{Fit}}, p_2^{\text{Fit}} \dots p_6^{\text{Fit}} : p_1^{\text{Fjt}}, p_2^{\text{Fjt}} \dots p_4^{\text{Fjt}} : p_1^{\text{Pat}}, \dots]^T$

The system then performs a principal component analysis on the shape vectors generated by all the training images to generate a shape appearance model for the mouth (defined by matrix M_s) which relates each mouth shape vector to a corresponding vector of shape mouth parameters by:

10 $\mathbf{p}_v^{\text{Ms}} = M_s (\mathbf{p}_v^{\text{FMs}} - \bar{\mathbf{p}}^{\text{FMs}}) \quad (5)$

15 where \mathbf{p}_v^{Ms} is the mouth shape vector for the mouth in the v -th training image, $\bar{\mathbf{p}}^{\text{FMs}}$ is the mean mouth shape vector from the training vectors and \mathbf{p}_v^{Ms} is a vector of mouth shape parameters for the mouth shape vector $\mathbf{p}_v^{\text{FMs}}$. The mouth shape model, defined by matrix M_s , describes the main modes of variation of the shape of the mouths within the training images; and the vector of mouth shape parameters (\mathbf{p}_v^{Ms}) for the mouth in the v -th training image has a parameter associated with each mode of variation whose value relates the shape of the input mouth to the corresponding mode of variation.

20 As with the facet appearance models, equation (5) above

can be rewritten with respect to the mouth shape vector p^{MS}_v to give:

$$p^{MS}_v = \bar{D}^{MS} - M_s^T D^{MS}_v \quad (6)$$

5

since $M_s M_s^T$ equals the identity matrix. Therefore, by modifying the mouth shape parameters, new mouth shapes can be generated which will be similar to those in the training set.

10

The system then performs a principal component analysis on the mouth texture parameter vectors (p^{Mt}) which are generated for the training images. This principal component analysis generates a mouth texture model (defined by matrix M_t) which relates each of the facet texture parameter vectors for the facets associated with the mouth, to a corresponding vector of mouth texture parameters, by:

15

$$D^{Mt}_v = M_t (p^{Mt}_v - \bar{D}^{Mt}) \quad (7)$$

20

where p^{Mt}_v is a vector of mouth facet texture parameters generated by the facet appearance models associated with the mouth for the mouth in the v -th training image; \bar{D}^{Mt} is the mean vector of mouth facet texture parameters from the training vectors and p^{Mt}_v is a vector of mouth texture parameters for the facet texture parameters p^{Mt}_v . The matrix M_t describes the main modes of variation within

the training images of the facet texture parameters generated by the facet appearance models which are associated with the mouth; and the vector of mouth texture parameters (p^{Mt}_v) has a parameter associated with each of those modes of variation whose value relates the texture of the input mouth to the corresponding mode of variation.

10 The processing then proceeds to step s71 shown in Figure 9 where the system determines the number of shape parameters and texture parameters needed to describe the training data received from the facet appearance models which are associated with the mouth. As shown in Figure 10, in this embodiment, the mouth appearance model 63 requires five shape parameters and four texture parameters to be able to model most of this variation. The system therefore stores the appropriate mouth shape and texture appearance model matrices for subsequent use.

20 As those skilled in the art will appreciate, a similar procedure is performed to determine each of the appearance models shown in Figure 5, starting from the facet appearance models at the base of the hierarchy. A further description of how these remaining appearance models are determined will, therefore, not be given here. The resulting hierarchical appearance model allows a small number of global face appearance parameters to be input to the face appearance model 61, which generates

further parameters which propagate down through the hierarchical model structure until facet pixel values are generated, from which an image which corresponds to the global appearance parameters can be generated.

5

AUTOMATIC GENERATION OF APPEARANCE PARAMETERS

In the description given above of the way in which the appearance models are generated, appearance parameters for an image were generated from a manual placement of a number of landmark points over the image. However, during use of the appearance model to track the first actor's head in the source video sequence and during the calculation of the difference parameters (D_{diff}), the appearance parameters for the heads in the input images were automatically calculated. This task involves finding the set of global appearance parameters p which best describe the pixels in view. This problem is complicated because the inverse of each of the appearance models in the hierarchical appearance model is not necessarily one-to-one. In this embodiment, the appearance parameters for the head in an input image are calculated in a two-step process. In the first step, an initial set of global appearance parameters for the head in the current frame (I_s^i) is found using a simple and rapid technique. For all but the first frame of the source video sequence, this is achieved by simply using the appearance parameters from the preceding video frame (I_s^{i-1}) before modification in step s3 (i.e. parameters 10 15 20 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 25510 25511 25512 25513 25514 25515 25516 25517 25518 25519 25520 25521 25522 25523 25524 25525 25526 25527 25528 25529 25530 25531 25532 25533 25534 25535 25536 25537 25538 25539 255310 255311 255312 255313 255314 255315 255316 255317 255318 255319 255320 255321 255322 255323 255324 255325 255326 255327 255328 255329 255330 255331 255332 255333 255334 255335 255336 255337 255338 255339 255340 255341 255342 255343 255344 255345 255346 255347 255348 255349 255350 255351 255352 255353 255354 255355 255356 255357 255358 255359 255360 255361 255362 255363 255364 255365 255366 255367 255368 255369 255370 255371 255372 255373 255374 255375 255376 255377 255378 255379 255380 255381 255382 255383 255384 255385 255386 255387 255388 255389 2553810 2553811 2553812 2553813 2553814 2553815 2553816 2553817 2553818 2553819 2553820 2553821 2553822 2553823 2553824 2553825 2553826 2553827 2553828 2553829 25538210 25538211 25538212 25538213 25538214 25538215 25538216 25538217 25538218 25538219 25538220 25538221 25538222 25538223 25538224 25538225 25538226 25538227 25538228 25538229 25538230 25538231 25538232 25538233 25538234 25538235 25538236 25538237 25538238 25538239 25538240 25538241 25538242 25538243 25538244 25538245 25538246 25538247 25538248 25538249 25538250 25538251 25538252 25538253 25538254 25538255 25538256 25538257 25538258 25538259 25538260 25538261 25538262 25538263 25538264 25538265 25538266 25538267 25538268 25538269 25538270 25538271 25538272 25538273 25538274 25538275 25538276 25538277 25538278 25538279 25538280 25538281 25538282 25538283 25538284 25538285 25538286 25538287 25538288 25538289 25538290 25538291 25538292 25538293 25538294 25538295 25538296 25538297 25538298 25538299 255382100 255382101 255382102 255382103 255382104 255382105 255382106 255382107 255382108 255382109 255382110 255382111 255382112 255382113 255382114 255382115 255382116 255382117 255382118 255382119 255382120 255382121 255382122 255382123 255382124 255382125 255382126 255382127 255382128 255382129 255382130 255382131 255382132 255382133 255382134 255382135 255382136 255382137 255382138 255382139 255382140 255382141 255382142 255382143 255382144 255382145 255382146 255382147 255382148 255382149 255382150 255382151 255382152 255382153 255382154 255382155 255382156 255382157 255382158 255382159 255382160 255382161 255382162 255382163 255382164 255382165 255382166 255382167 255382168 255382169 255382170 255382171 255382172 255382173 255382174 255382175 255382176 255382177 255382178 255382179 255382180 255382181 255382182 255382183 255382184 255382185 255382186 255382187 255382188 255382189 255382190 255382191 255382192 255382193 255382194 255382195 255382196 255382197 255382198 255382199 255382200 255382201 255382202 255382203 255382204 255382205 255382206 255382207 255382208 255382209 255382210 255382211 255382212 255382213 255382214 255382215 255382216 255382217 255382218 255382219 255382220 255382221 255382222 255382223 255382224 255382225 255382226 255382227 255382228 255382229 255382230 255382231 255382232 255382233 255382234 255382235 255382236 255382237 255382238 255382239 255382240 255382241 255382242 255382243 255382244 255382245 255382246 255382247 255382248 255382249 255382250 255382251 255382252 255382253 255382254 255382255 255382256 255382257 255382258 255382259 255382260 255382261 255382262 255382263 255382264 255382265 255382266 255382267 255382268 255382269 255382270 255382271 255382272 255382273 255382274 255382275 255382276 255382277 255382278 255382279 255382280 255382281 255382282 255382283 255382284 255382285 255382286 255382287 255382288 255382289 255382290 255382291 255382292 255382293 255382294 255382295 255382296 255382297 255382298 255382299 255382300 255382301 255382302 255382303 255382304 255382305 255382306 255382307 255382308 255382309 255382310 255382311 255382312 255382313 255382314 255382315 255382316 255382317 255382318 255382319 255382320 255382321 255382322 255382323 255382324 255382325 255382326 255382327 255382328 255382329 255382330 255382331 255382332 255382333 255382334 255382335 255382336 255382337 255382338 255382339 255382340 255382341 255382342 255382343 255382344 255382345 255382346 255382347 255382348 255382349 255382350 255382351 255382352 255382353 255382354 255382355 255382356 255382357 255382358 255382359 255382360 255382361 255382362 255382363 255382364 255382365 255382366 255382367 255382368 255382369 255382370 255382371 255382372 255382373 255382374 255382375 255382376 255382377 255382378 255382379 255382380 255382381 255382382 255382383 255382384 255382385 255382386 255382387 255382388 255382389 255382390 255382391 255382392 255382393 255382394 255382395 255382396 255382397 255382398 255382399 255382400 255382401 255382402 255382403 255382404 255382405 255382406 255382407 255382408 255382409 255382410 255382411 255382412 255382413 255382414 255382415 255382416 255382417 255382418 255382419 255382420 255382421 255382422 255382423 255382424 255382425 255382426 255382427 255382428 255382429 255382430 255382431 255382432 255382433 255382434 255382435 255382436 255382437 255382438 255382439 255382440 255382441 255382442 255382443 255382444 255382445 255382446 255382447 255382448 255382449 255382450 255382451 255382452 255382453 255382454 255382455 255382456 255382457 255382458 255382459 255382460 255382461 255382462 255382463 255382464 255382465 255382466 255382467 255382468 255382469 255382470 255382471 255382472 255382473 255382474 255382475 255382476 255382477 255382478 255382479 255382480 255382481 255382482 255382483 255382484 255382485 255382486 255382487 255382488 255382489 255382490 255382491 255382492 255382493 255382494 255382495 255382496 255382497 255382498 255382499 255382500 255382501 255382502 255382503 255382504 255382505 255382506 255382507 255382508 255382509 255382510 255382511 255382512 255382513 255382514 255382515 255382516 255382517 255382518 255382519 255382520 255382521 255382522 255382523 255382524 255382525 255382526 255382527 255382528 255382529 255382530 255382531 255382532 255382533 255382534 255382535 255382536 255382537 255382538 255382539 255382540 255382541 255382542 255382543 255382544 255382545 255382546 255382547 255382548 255382549 255382550 255382551 255382552 255382553 255382554 255382555 255382556 255382557 255382558 255382559 255382560 255382561 255382562 255382563 255382564 255382565 255382566 255382567 255382568 255382569 255382570 255382571 255382572 255382573 255382574 255382575 255382576 255382577 255382578 255382579 255382580 255382581 255382582 255382583 255382584 255382585 255382586 255382587 255382588 255382589 255382590 255382591 255382592 255382593 255382594 255382595 255382596 255382597 255382598 255382599 255382600 255382601 255382602 255382603 255382604 255382605 255382606 255382607 255382608 255382609 255382610 255382611 255382612 255382613 255382614 255382615 255382616 255382617 255382618 255382619 255382620 255382621 255382622 255382623 255382624 255382625 255382626 255382627 255382628 255382629 255382630 255382631 255382632 255382633 255382634 255382635 255382636 255382637 255382638 255382639 255382640 255382641 255382642 255382643 255382644 255382645 255382646 255382647 255382648 255382649 255382650 255382651 255382652 255382653 255382654 255382655 255382656 255382657 255382658 255382659 255382660 255382661 255382662 255382663 255382664 255382665 255382666 255382667 255382668 255382669 255382670 255382671 255382672 255382673 255382674 255382675 255382676 255382677 255382678 255382679 255382680 255382681 255382682 255382683 255382684 255382685 255382686 255382687 255382688 255382689 255382690 255382691 255382692 255382693 255382694 255382695 255382696 255382697 255382698 255382699 255382700 255382701 255382702 255382703 255382704 255382705 255382706 255382707 255382708 255382709 255382710 255382711 255382712 255382713 255382714 255382715 255382716 255382717 255382718 255382719 255382720 255382721 255382722 255382723 255382724 255382725 255382726 255382727 255382728 255382729 255382730 255382731 255382732 255382733 255382734 255382735 255382736 255382737 255382738 255382739 255382740 255382741 255382742 255382743 255382744 255382745 255382746 255382747 255382748 255382749 255382750 255382751 255382752 255382753 255382754 255382755 255382756 255382757 255382758 255382759 255382760 255382761 255382762 255382763 255382764 255382765 255382766 255382767 255382768 255382769 255382770 255382771 255382772 255382773 255382774 255382775 255382776 255382777 255382778 255382779 255382780 255382781 255382782 255382783 255382784 255382785 255382786 255382787 255382788 255382789 255382790 255382791 255382792 255382793 255382794 255382795 255382796 255382797 255382798 255382799 255382800 255382801 255382802 255382803 255382804 255382805 255382806 255382807 255382808 255382809 255382810 255382811 255382812 255382813 255382814 255382815 255382816 255382817 255382818 255382819 255382820 255382821 255382822 255382823 255382824 255382825 255382826 255382827 255382828 255382829 255382830 255382831 255382832 255382833 255382834 255382835 255382836 255382837 255382838 255382839 255382840 255382841 255382842 255382843 255382844 255382845 255382846 255382847 255382848 255382849 255382850 255382851 255382852 255382853 255382854 255382855 255382856 255382857 255382858 255382859 255382860 255382861 255382862 255382863 255382864 255382865 255382866 255382867 255382868 255382869 255382870 255382871 255382872 255382873 255382874 255382875 255382876 255382877 255382878 255382879 255382880 255382881 255382882 255382883 255382884 255382885 255382886 255382887 255382888 255382889 255382890 255382891 255382892 255382893 255382894 255382895 255382896 255382897 255382898 255382899 255382900 255382901 255382902

p_s^{i-1}). In this embodiment, the global appearance parameters (p) effectively define the shape and colour texture of the head. For the first frame and for the target image the initial estimate of the appearance parameters is set to the mean set of appearance parameters and the scale, position and orientation is initially estimated by the user manually placing the mean head over the head in the image.

10 In the second step, an iterative technique is used in order to make fine adjustments to the initial estimate of the appearance parameters. The adjustments are made in an attempt to minimise the difference between the head described by the global appearance parameters (the model head) and the head in the current video frame (the image head). With 50 appearance parameters, this represents a difficult optimisation problem. This can be performed by using a standard steepest descent optimisation technique to iteratively reduce the mean squared error between the given image pixels and those predicted by a particular set of appearance parameter values. In particular, minimising the following error function $E(p)$:

$$E(p) = [I^a - F(p)]^T [I^a - F(p)] \quad (8)$$

where I^a is a vector of actual image RGB pixel values at the locations where the appearance model predicts values

(the appearance model does not predict all pixel values since it ignores background pixels and only predicts a subsample of pixel values within the object being modelled) and $F(p)$ is the vector of image RGB pixel values predicted by the hierarchical appearance model.

5 As those skilled in the art will appreciate, $E(p)$ will only be zero when the model head (i.e. $F(p)$) predicts the actual image head (I^a) exactly. Standard steepest descent optimisation techniques stipulate that a step in the direction $-\nabla E(p)$ should result in a reduction in the error function $E(p)$, provided the error function is well behaved. Therefore, the change (Δp) in the set of parameter values should be:

10

$$15 \quad \Delta \bar{p} = 2[\nabla F(p)]^T[I^a - F(p)] \quad (9)$$

which requires the calculation of the differential of the appearance model, i.e. $\nabla F(p)$.

20 The technique described by Edwards et al assumes that, on average over the whole parameter space, $\nabla F(p)$ is constant. The update equation then becomes:

$$25 \quad \Delta p = A[I^a - F(p)] \quad (10)$$

for some constant matrix A (referred to as the "Active matrix") which is determined beforehand during a training routine. In this embodiment, rather than using a single

constant matrix associated with the entire hierarchical appearance model, an Active matrix is determined and used for each of the individual appearance models which form part of the hierarchical appearance model. The way in which these Active matrices are determined in this embodiment will now be described with reference to Figures 11a and 11b, which illustrate the processing steps performed to generate the Active matrix for each facet appearance model and the Active matrix for the mouth appearance model.

As shown in Figure 11a, in step s73, the system chooses a random facet parameter vector (p^{F_i}) for the current facet (i) and then, in s75, perturbs this facet parameter vector by a small random amount to create $p^{F_i} + \Delta p^{F_i}$. In this embodiment, the facet parameter vectors include not only the texture parameters, but also the six shape parameters which define the (x,y) coordinates of the facet's location within the image. The processing then proceeds to step s77 where the system uses the parameter vector p^{F_i} and the perturbed parameter vector $p^{F_i} + \Delta p^{F_i}$ to create model images $I_0^{F_i}$ and $I_1^{F_i}$ respectively. The processing then proceeds to step s79 where the system records the parameter change Δp^{F_i} and image difference $I_1^{F_i} - I_0^{F_i}$. Then in step s81, the system determines whether or not there is sufficient training data for the current facet. If there is not then the processing returns to step s21. Once sufficient training data has been

generated, the processing proceeds to step s83 where the system performs multiple multivariate linear regressions on the data for the current facet to identify an Active matrix (A_{fi}) for the current facet.

5

Figure 11b shows the processing steps required to calculate the Active matrix for the mouth appearance model. As shown, in step s85, the system chooses a random mouth parameter vector p^M . In this embodiment, this vector includes both the mouth shape parameters and the mouth texture parameters. Then, in step s87, the system perturbs this mouth parameter vector by a small random amount to create $p^M + \Delta p^M$. The processing then proceeds to step s89 where the system uses the mouth parameter vectors p^M and the perturbed mouth parameter vector $p^M + \Delta p^M$ to create model images I_0^M and I_1^M respectively, using the mouth appearance model and the facet appearance models associated with the mouth. The processing then proceeds to step s91 where the facet appearance models associated with the mouth are used again to transform the mouth model images I_0^M and I_1^M into corresponding facet appearance parameters p_0^{FM} and p_1^{FM} , which are then subtracted to determine the corresponding change Δp^{FM} in the mouth facet parameters. The processing then proceeds to step s93 where the system records the mouth parameter change Δp^M and the mouth facet parameter change Δp^{FM} . The processing then proceeds to step s95 where the system determines whether

or not there is sufficient training data. If there is not, then the processing returns to step s85. Once sufficient training data has been generated, the processing proceeds to step s97, where the system 5 performs multiple multivariate linear regressions on the training data for the mouth to identify the Active matrix (A_m) for the mouth which relates changes in mouth parameters Δp^m to changes in facet parameters Δp^{fm} for the facets associated with the mouth.

10

As those skilled in the art will appreciate, a similar processing technique is used in order to identify the Active matrix for each of the appearance models shown in Figure 5.

15

Once the Active matrices have been determined for the hierarchical appearance model, they can then be used to iteratively update a current estimate of a set of appearance parameters for an input image. Figure 12 20 illustrates the processing steps performed in this iterative routine for the current source video frame. As shown, in step s101, the system initially estimates a set of global parameters for the head in the current source video frame. The processing then proceeds to step s103 25 where the system generates a model image from the estimated global parameters and the hierarchical appearance model. The system then proceeds to step s105 where it determines the image error between the model

image and the current source video frame. Then, in step s107, the system uses this image error to propagate parameter changes up the hierarchy of the hierarchical appearance model using the stored Active matrices to determine a change in the global parameters. This change in global parameters is then used, in step s109, to update the current global parameters for the current source video frame. The system then determines, in step s111, whether or not convergence has been reached by comparing the error obtained from equation (8) using the updated global parameters with a predetermined threshold (Th). If convergence has not been reached, then the processing returns to step s103. Once convergence is reached, the processing proceeds to step s113, where the current global appearance parameters are output as the global appearance parameters for the current source video frame and then the processing ends.

ALTERNATIVE EMBODIMENTS

20 In the above embodiment, the same hierarchical model structure was used to model the variation in the shape and texture within the training images. As those skilled in the art will appreciate, one model hierarchy can be used to model the shape variation and a different model hierarchy can be used to model the texture variation. Alternatively still, rather than separating the shape and texture parameters, each of the appearance models within the hierarchical model may model the combined variation

of the shape and texture within the training images.

In the above embodiments, a facet appearance model was generated for each facet defined within the training images. As those skilled in the art will appreciate, many of the facets may be grouped together such that a single facet appearance model is generated for those facets. In one form of such an embodiment, a single facet appearance model may be determined which models the variability of texture within each facet of the training images.

In the above embodiments, the same amount of texture information was extracted from each facet within the training images. In particular, fifty RGB texture values were extracted from each training facet. In an alternative embodiment, the amount of texture information extracted from each facet may vary in dependence upon the size of the facet. For example, more texture information may be extracted from larger facets or more texture information may be extracted from facets associated with important features of the face, such as the mouth, eyes or nose.

In the above embodiments, each appearance model was determined from a principal component analysis of a set of training data. This principal component analysis determines a linear relationship between the training

data and a set of model parameters. As those skilled in the art will appreciate, techniques other than principal component analysis can be used to determine a parametric model which relates a set of parameters to the training data. This model may define a non-linear relationship between the training data and the model parameters. For example, one or more of the models within the hierarchy may comprise a neural network which relates the set of input parameters to the training data.

10

In the above embodiments, a principal component analysis was performed on a set of training data in order to identify a relatively small number of parameters which describe the main modes of variation within the training data. This allows a relatively small number of input parameters to be able to generate a larger set of output parameters from the model. However, as those skilled in the art will appreciate, this is not essential. One or more of the appearance models may act as transformation models in which the number input parameters is the same as or greater than the number of output parameters. This can be used to generate a set of input parameters which can be changed by the user in some intuitive way. For example, in order to identify parameters which have a linear relationship with features in the object, such as a parameter that linearly changes the amount of smile within a face image.

In the above embodiments, a set of Active matrices were used in order to identify automatically a set of appearance parameters for an input image. As those skilled in the art will appreciate, rather than having separate Active matrices for each of the components in the hierarchical appearance model, a global Active matrix may be used instead. Further, although both the shape and grey level parameters were used in order to derive the Active matrices, suitable Active matrices can be determined using just the shape information.

In the above embodiments, the variation in both the shape and texture within the training images were modelled. As those skilled in the art will appreciate, this hierarchical modelling technique can be used to model only the shape of the objects within the training images. Such a shape model could be then used to track objects within a video sequence.

In the first embodiment, the target image illustrated a computer generated head. This is not essential. For example, the target image might be a hand-drawn head or an image of a real person. Figures 13d and 13e illustrate how an embodiment with a hand-drawn character might be used in character animation. In particular, Figure 13d shows a hand-drawn sketch of a character which, when combined with the images from the source video sequence (some of which are shown in Figure 13a)

generate a target video sequence, some frames of which are shown in Figure 13e. As can be seen from a comparison of the corresponding frames in the source and target video frames, the hand-drawn sketch has been 5 animated automatically using this technique. As those skilled in the art will appreciate, this is a much quicker and simpler technique for achieving computer animation as compared with existing systems which require the animator to manually create each frame of the 10 animation. In particular, in this embodiment, all that is required is a video sequence of a real life actor acting out the scene to be animated, together with a single sketch of the character to be animated.

15 The above embodiment has described the way in which a target image can be used to modify a source video sequence. In order to do this, a set of appearance parameters has to be automatically calculated for each frame in the video sequence. This involved the use of a 20 number of Active matrices which relate image errors to appearance parameter changes. As those skilled in the art will appreciate, similar processing is required in other applications, such as the tracking of an object 25 within a video sequence, the tracking of a human face within a video sequence or the tracking of a knee joint in an MRI scan.

In the above embodiment, the appearance model was used to

model the variations in facial expressions and 3D pose of human heads. As those skilled in the art will appreciate, the appearance model can be used to model the appearance of any deformable object such as parts of the body and other animals and objects. For example, the above techniques can be used to track the movement of lips in a video sequence. Such an embodiment could be used in film dubbing applications in order to synchronise the lip movements with the dubbed sound. This animation technique might also be used to give animals and other objects human-like characteristics by combining images of them with a video sequence of an actor. This technique can also be used for monitoring the shape and appearance of objects passing along a production line for quality control purposes.

In the above embodiment, the appearance model was generated by using a principal component analysis of shape and texture data which is extracted from the training images. As those skilled in the art will appreciate, by modelling the features of the training heads in this way, it is possible to accurately model each head by just a small number of parameters. However, other modelling techniques, such as vector quantisation and wavelet techniques can be used.

In the above embodiments, the training images used to generate the appearance model were all colour images in

which each pixel had an RGB value. As those skilled in the art will appreciate, the way in which the colour is represented in this embodiment is not important. In particular, rather than each pixel having a red, green and blue value, they might be represented by a chrominance and a luminance component or by hue, saturation and value components. Alternatively still, the training images may be black and white images, in which case only grey level data would be extracted from the facets in the training images. Additionally, the resolution of each training image may be different.

In the above embodiment, during the automatic generation of the appearance parameters, and in particular during the iterative updating of these appearance parameters the error between the input image and the model image was generated using the appearance model. Since this iterative technique still requires a relatively accurate initial estimate for the appearance parameters, it is possible initially to perform the iterations using lower resolution images and once convergence has been reached for the lower resolutions to then increase the resolution of the images and to repeat the iterations for the higher resolutions. In such an embodiment, separate Active 25 matrices would be required for each of the resolutions.

In the above embodiment, the difference parameters were determined by comparing the image of the first actor from

one of the frames of the source video sequence with the image of the second actor in the target image. In an alternative embodiment, a separate image of the first actor may be provided which does not form part of the source video sequence.

In the above embodiments, each of the appearance models modelled variations in two-dimensional images. The above modelling technique could be adapted to work with 3D images and animations. In such an embodiment, the training images used to generate the appearance model would normally include 3D images instead of 2D images. The three-dimensional models may be obtained using a three dimensional scanner which typically work either by using laser range-finding over the object or by using one or more stereo pairs of cameras. Once a 3D hierarchical appearance model has been created from the training models, new 3D models can be generated by adjusting the appearance parameters and existing 3D models can be animated using the same differencing technique that was used in the two-dimensional embodiment described above. This 3D model can then be used to track 3D objects directly within a 3D animation. Alternatively, a 2D model may be used to track the 3D object within a video sequence and then use the result to generate 3D data for the tracked object.

In the above embodiment, a set of difference parameters

were identified which describe the main differences between the head in the video sequence and the head in the target image, which difference parameters were used to modify the video sequence so as to generate a target video sequence showing the second head. In the embodiment, the set of difference parameters were added to a set of appearance parameters for the current frame being processed. In an alternative embodiment, the difference parameters may be weighted so that, for example, the target video sequence shows a head having characteristics from both the first and second actors.

In the above embodiment, a hierarchical appearance model is used to model the appearance of human faces. The model is then used to modify a source video sequence showing a first actor performing a scene to generate a target video sequence showing a second actor performing the same scene. As those skilled in the art will appreciate, the hierarchical model presented above can be used in various other applications. For example, the hierarchical appearance model can be used for synthetic two-dimensional or three-dimensional character generation; video compression when the video is substantially that of an object which is modelled by the appearance model; object recognition for security purposes; face tracking for human performance analysis or human computer interaction and the like; 3D model generation from two-dimensional images; and image editing

(for example making people look older or younger, fatter or thinner etc).

5 In the above embodiment, an iterative process was used to update an estimated set of appearance parameters for an input image. This iterative process continued until an error between the actual image and the image predicted by the model was below a predetermined threshold. In an alternative embodiment, where there is only a predetermined amount of time available for determining a set of appearance parameters for an input image, this iterative routine may be performed for a predetermined period of time or for a predetermined number of iterations.

10

CLAIMS:

1. A parametric model for modelling the shape of an object, the model comprising:

5 data defining a function which relates a set of input parameters to a set of locations which identify the relative positions of a plurality of predetermined points on the object;

10 characterised in that said data defines a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from 15 one or more functions from a higher layer of the hierarchy and to generate therefrom at least some of said locations which identify the relative positions of said predetermined points.

20 2. A model according to claim 1, wherein said hierarchy comprises one or more intermediate layers of functions which are operable to receive parameters output from one or more functions from a higher layer of the hierarchy and to generate therefrom a set of output parameters for 25 input to functions in a lower layer of the hierarchy.

3. A model according to claim 1 or 2, for modelling the two-dimensional shape of the object by identifying the

relative positions of said predetermined points in a predetermined plane.

4. A model according to claim 1 or 2, for modelling the
5 three-dimensional shape of the object by identifying the relative positions of the predetermined points in a three-dimensional space.

10 5. A model according to any preceding claim, wherein one or more of said functions comprises a linear function which linearly relates the input parameters to the function to the output parameters of the function.

15 6. A model according to claim 5, wherein said one or more linear functions are identified from a principal component analysis of training data derived from a set of training objects.

20 7. A model according to any preceding claim, wherein one or more of said functions are non-linear.

8. A model according to claim 7, wherein at least one of said non-linear functions comprises a neural network.

25 9. A model according to any preceding claim, wherein the number of parameters input to at least one of said functions is smaller than the number of parameters output from the function.

10. A model according to any preceding claim, wherein the number of input parameters to at least one of said functions is greater than or equal to the number of parameters output by the function.

5

11. A model according to any preceding claim for modelling the shape and texture of the object, the model further comprising data defining a hierarchical set of functions in which a function in a top layer of the hierarchy is operable to generate a set of output parameters from a set of input parameters and in which one or more functions in a bottom layer of the hierarchy are operable to receive parameters output from one or more functions from a higher layer of the hierarchy and to generate therefrom texture information for the object.

10

15

12. A model according to claim 11, wherein the texture hierarchy has the same structure as the shape hierarchy.

20

13. A model according to claim 11 or 12, wherein one or more of said functions are operable to relate an input set of shape and texture parameters to an output set of appearance parameters defining both shape and texture.

25

14. A model according to any preceding claim, wherein said object is a deformable object.

15. A model according to claim 14, wherein said

deformable object includes a human face.

16. A model according to claim 15, wherein said function in said top layer of the hierarchy models the shape of 5 the entire face and wherein said hierarchy includes a function which models the shape of the mouth.

17. A model according to claim 16, wherein said hierarchy further comprises a function for modelling the 10 shape of the eyes.

18. A model according to any preceding claim, wherein 15 the or each function in the bottom layer of the hierarchy identifies the positions of a plurality of predetermined points according to a predefined function of smaller number of control point positions.

19. A model according to claim 18, wherein the predefined function for each of the plurality of points 20 is a linear mapping of the control point positions and the control points are the three corners of a triangular facet.

25 20. A model according to claim 18, wherein the predefined function for each of the plurality of points is a predefined non-linear mapping of a fixed number of control point positions.

21. A model according to claim 18, wherein the predefined function for each of the plurality of points is a predefined displacement from a single control point.

5 22. A method of determining a set of appearance parameters representative of the appearance of an object, the method comprising the steps of:

10 (i) storing a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object;

15 (ii) storing at least one function which relates a change in the input parameters to an error between actual appearance data for the object and appearance data determined from the set of input parameters and said parametric model;

(iii) initially estimating a current set of input parameters for the object;

20 (iv) determining appearance data for the object from the current set of input parameters and the stored parametric model;

(v) determining the error between actual appearance data of the object and the appearance data determined from the current set of input parameters;

25 (vi) determining a change in the input parameters using said at least one stored function and said determined error; and

(vii) updating the current set of input parameters

with the determined change in the input parameters.

23. A method according to claim 22, further comprising the step of repeating steps (iv) to (vii) until the error determined in step (v) is less than a predetermined threshold.

10 24. A method according to claim 22, further comprising the step of repeating steps (iv) to (vii) for a predetermined amount of time or for a predetermined number of repetitions.

15 25. A method according to claim 22, 23 or 24, wherein said second storing step stores a plurality of functions, one associated with each function within the hierarchical model.

26. A method of tracking an object comprising the steps of:

20 (i) storing a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object;

25 (ii) storing at least one function which relates a change in the input parameters to an error between the actual appearance data for the object and the appearance data determined from the set of input parameters and said parametric model;

(iii) initially estimating a current set of input parameters for the object;

(iv) determining the appearance data for the object from the current set of input parameters and the stored 5 parametric model;

(v) determining an error between the actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

10 (vi) determining a change in the input parameters using the at least one stored function and the determined error;

(vii) updating the current set of input parameters with said change in the input parameters;

15 (viii) repeating steps (iv) to (vii) in order to reduce the error determined in step (v); and

(ix) repeating steps (iii) to (viii) to track the object.

20 27. An apparatus for determining a set of appearance parameters representative of the appearance of an object, the apparatus comprising:

means for storing (i) a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object; and (ii) at least one function which relates a change in the input parameters to an error between actual appearance data for the object and

the appearance data for the object determined from the set of input parameters and said parametric model;

means for receiving an initial estimate of a current set of input parameters for the object;

5 means for updating the current set of input parameters comprising:

(i) means for determining appearance data for the object from the current set of input parameters and the stored parametric model;

10 (ii) means for determining the error between the actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

15 (iii) means for determining a change in the input parameters using said at least one stored function and said determined error; and

(iv) means for updating the current set of input parameters with the determined change in the input parameters.

20

28. An apparatus according to claim 27, wherein said updating means is operable to update iteratively the current set of input parameters until the error determining means determines an error which is less than 25 a predetermined threshold.

29. An apparatus according to claim 27 or 28, wherein said storing means stores a plurality of functions, one

associated with each function within the hierarchical model.

30. An apparatus for tracking an object comprising:

5 means for storing (i) a parametric model according to any of claims 1 to 21 which relates a set of input parameters to appearance data representative of the appearance of the object; and (ii) at least one function which relates a change in the input parameters to an error between actual appearance data for the object and the appearance data for the object determined from the set of input parameters and said parametric model;

10 means for receiving an initial estimate of a current set of input parameters for the object;

15 means for updating the current set of input parameters comprising:

(i) means for determining appearance data for the object from the current set of input parameters and the stored parametric model;

20 (ii) means for determining an error between actual appearance data for the object and the appearance data for the object determined from the current set of input parameters;

25 (iii) means for determining a change in the input parameters using the at least one stored function and the determined error; and

(iv) means for updating the current set of input parameters with said change in the input parameters;

wherein said updating means is operable to update iteratively the current set of input parameters in order to reduce the determined error, wherein said receiving means is operable to receive further estimates of the 5 current input parameters and wherein said update means is operable to update the received estimates of the current input parameters in order to track said object.

31. A storage medium storing the parametric model 10 according to any of claims 1 to 21 or storing processor implementable instructions for controlling a processor to implement the method of any one of claims 22 to 26.

32. Processor implementable instructions for controlling 15 a processor to implement the method of any one of claims 22 to 26.

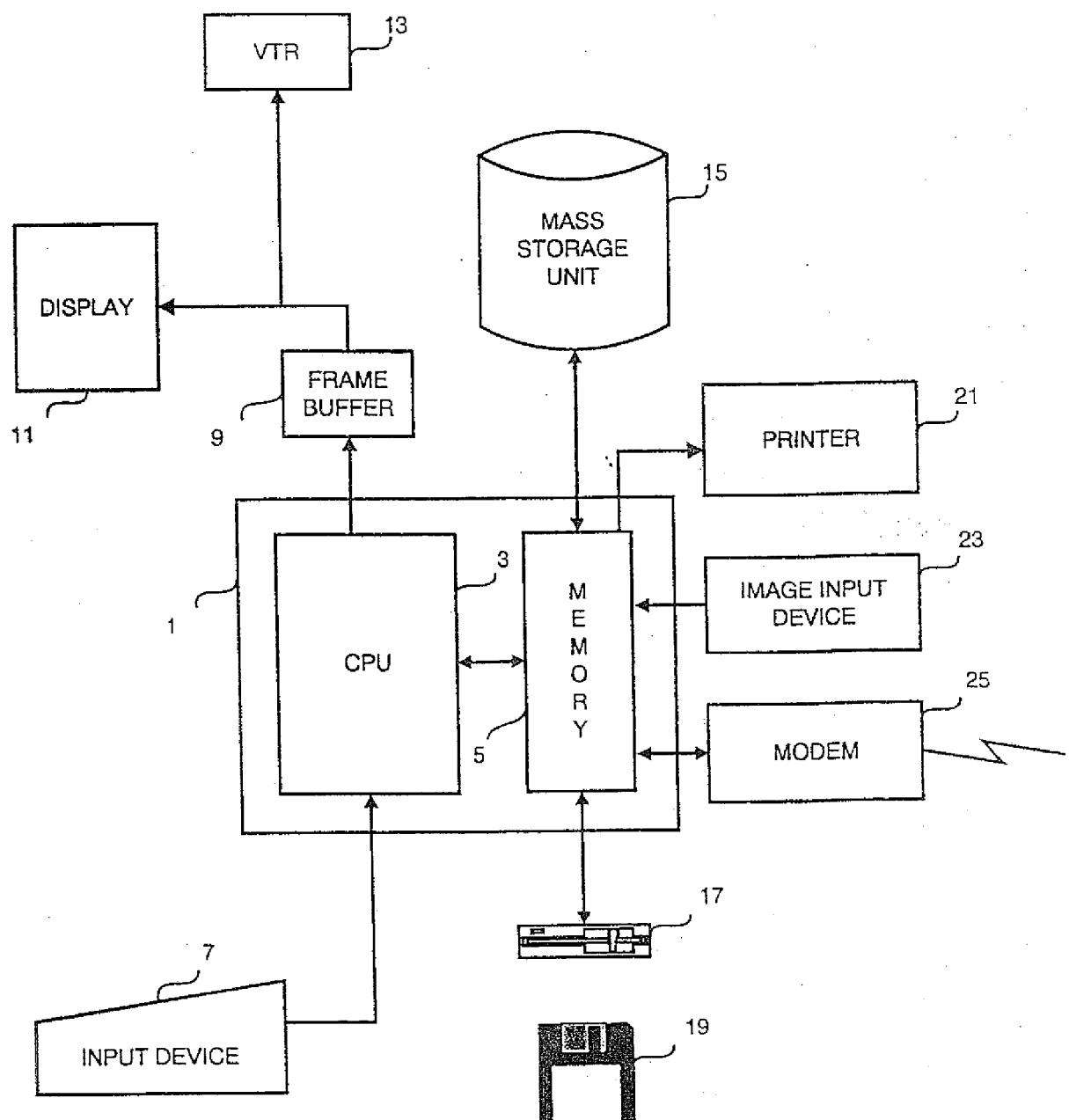


Fig. 1

2/14

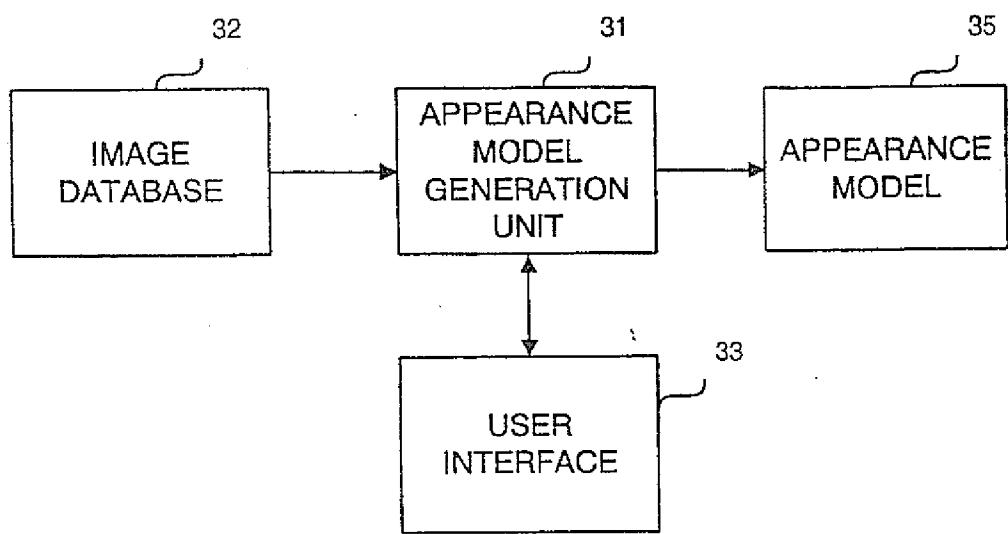


Fig. 2

3/14

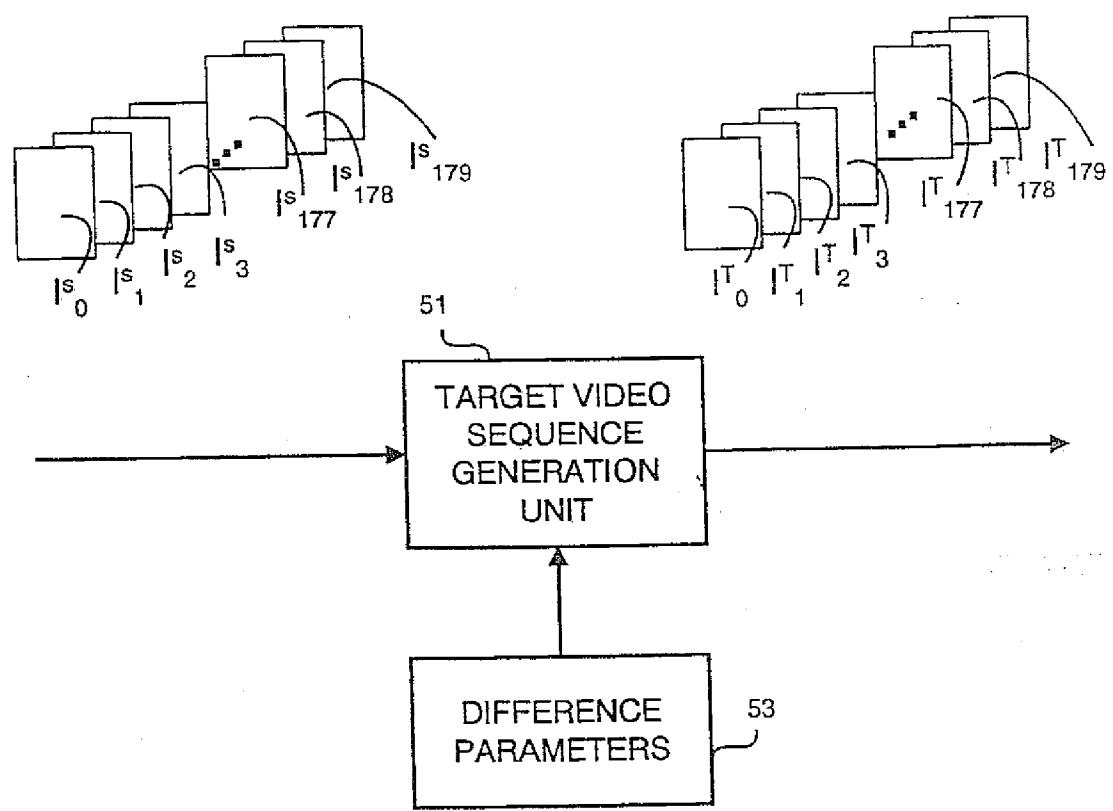


Fig. 3

4/14

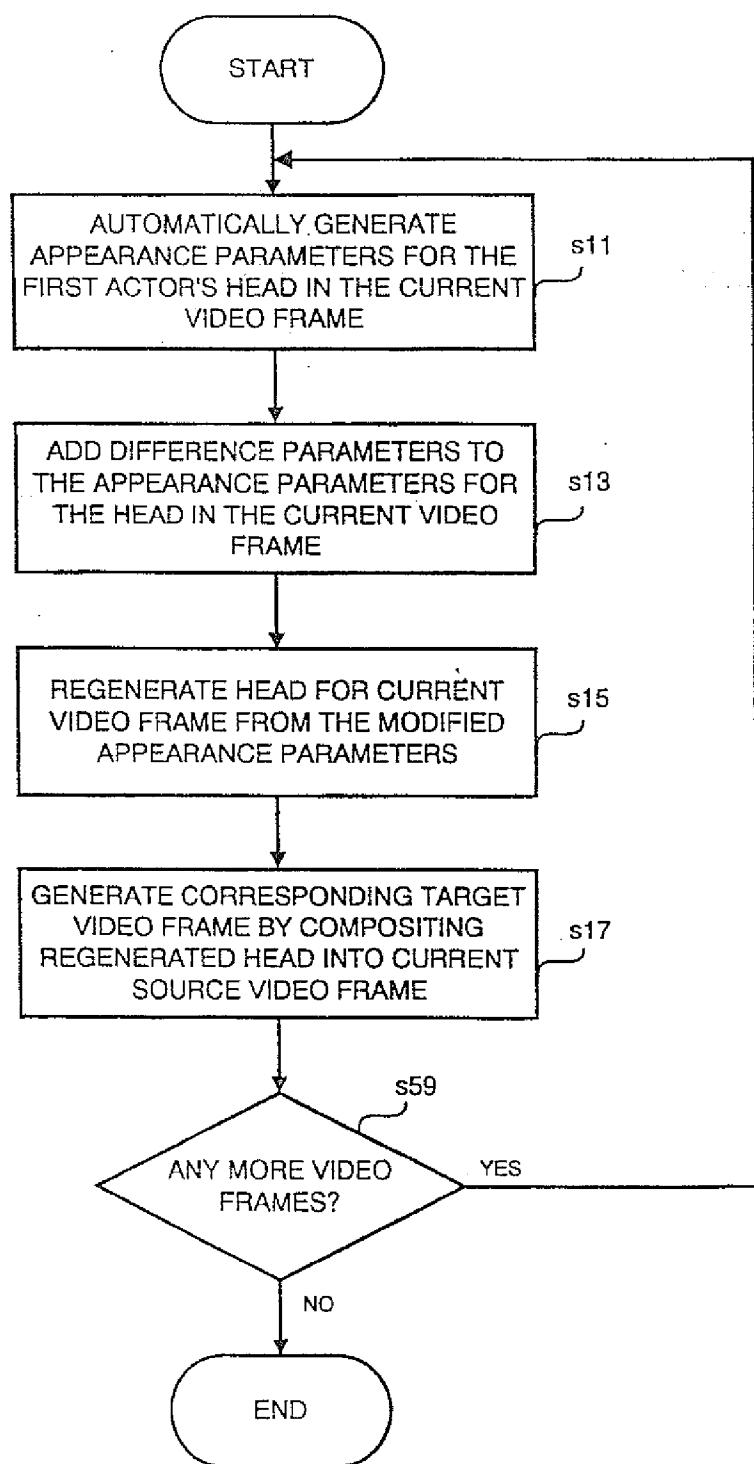


Fig. 4

5/14

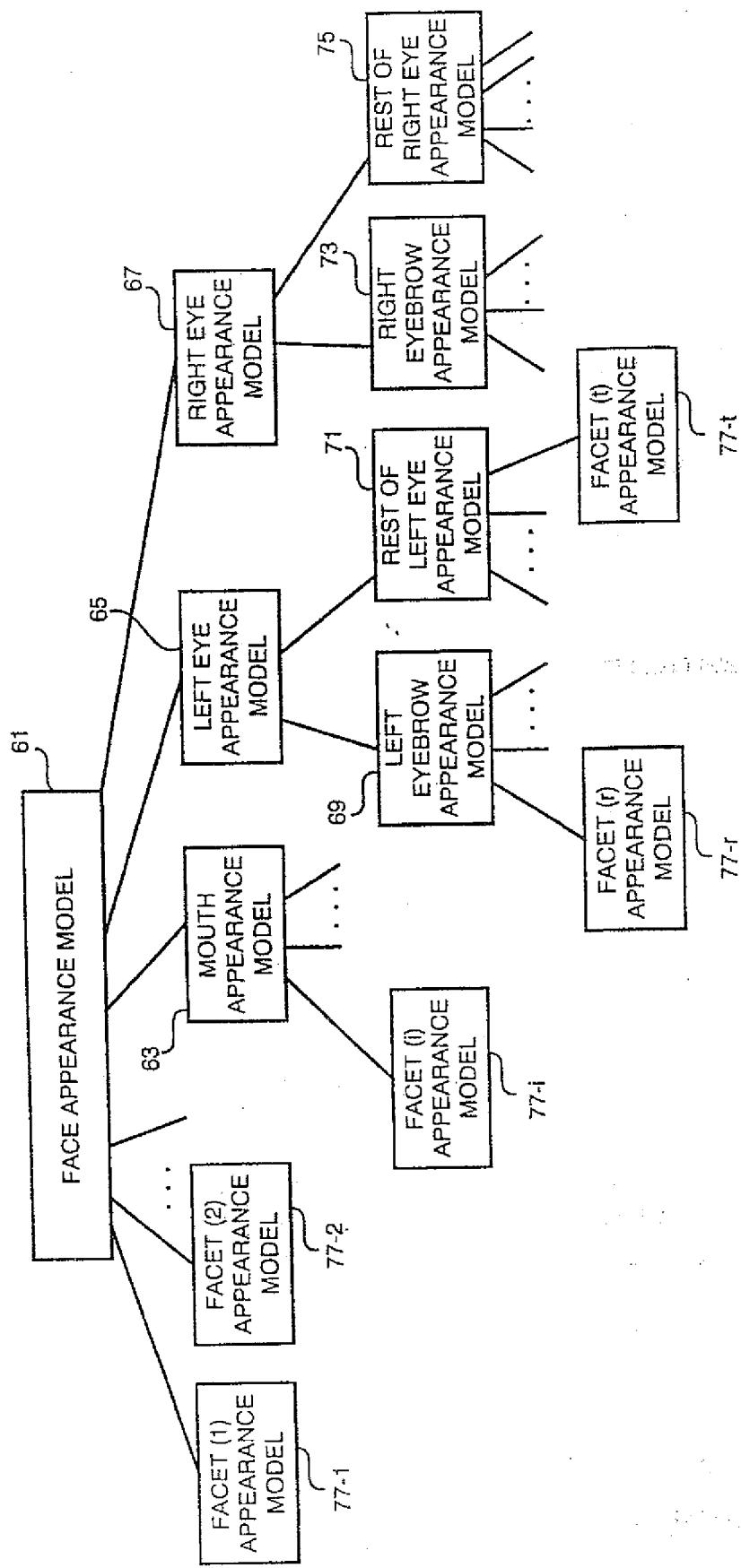


Fig. 5

6/14

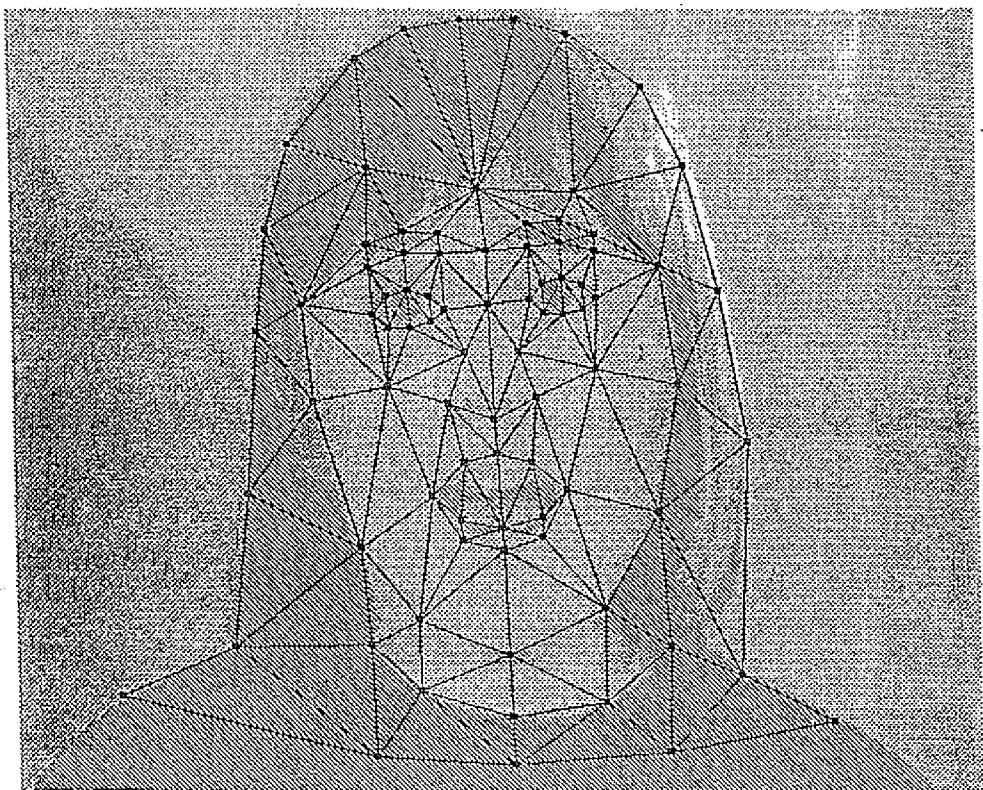


Fig. 6

7/14

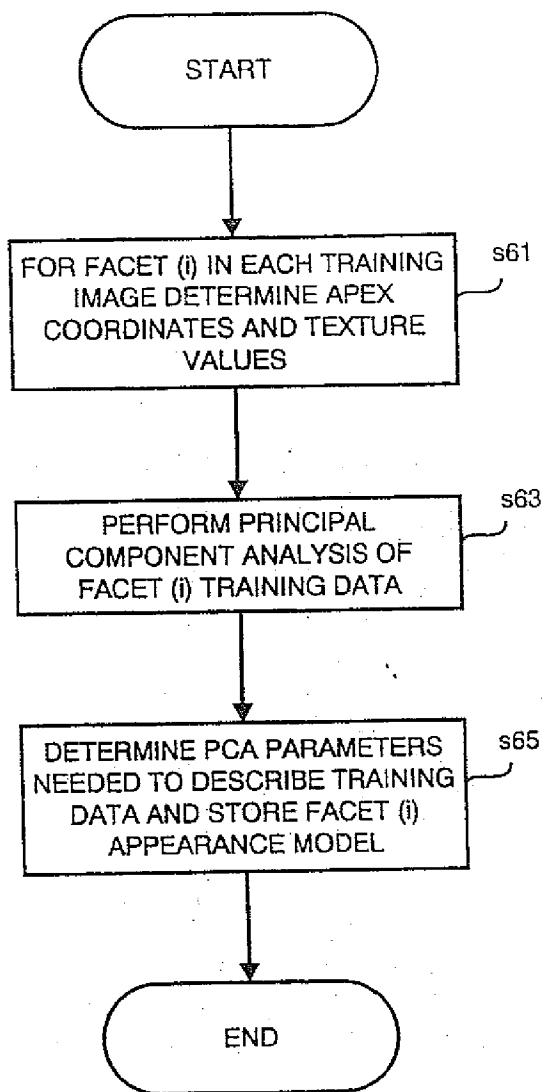


Fig. 7

8/14

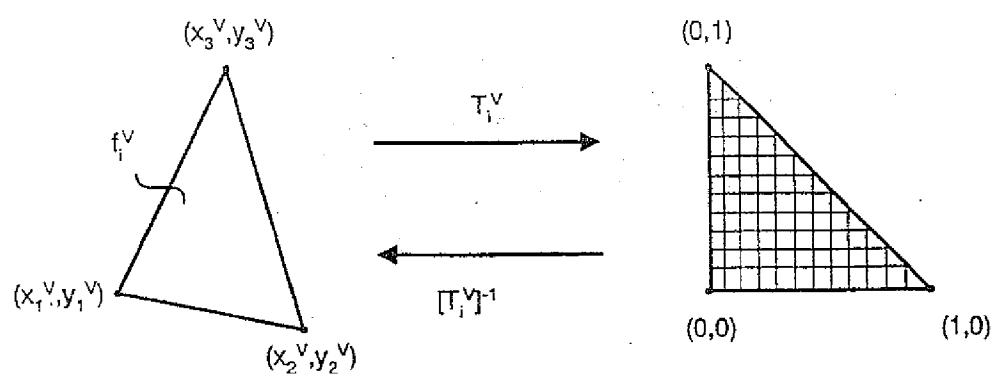


Fig. 8

9/14

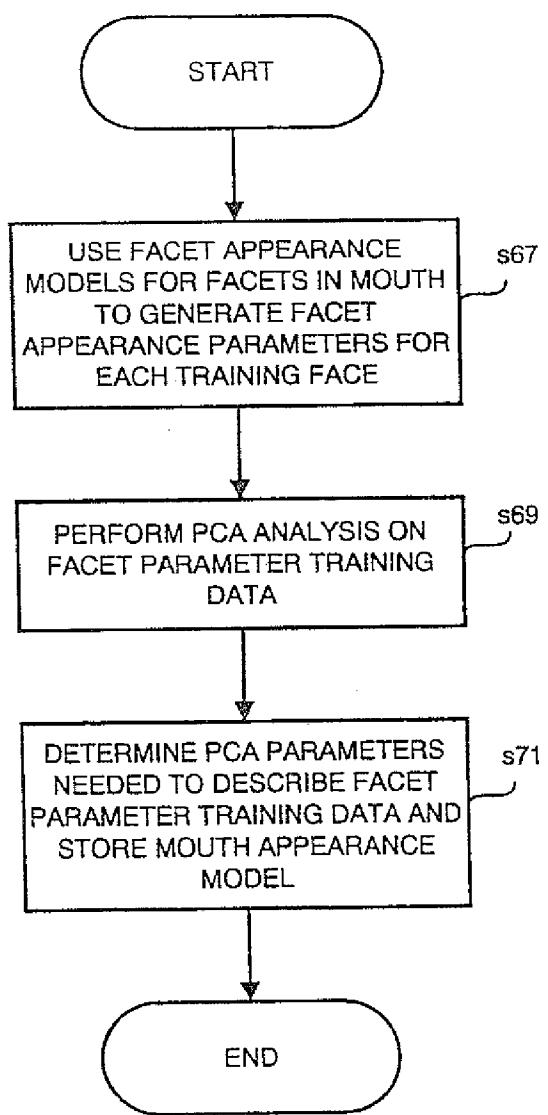


Fig. 9

10/14

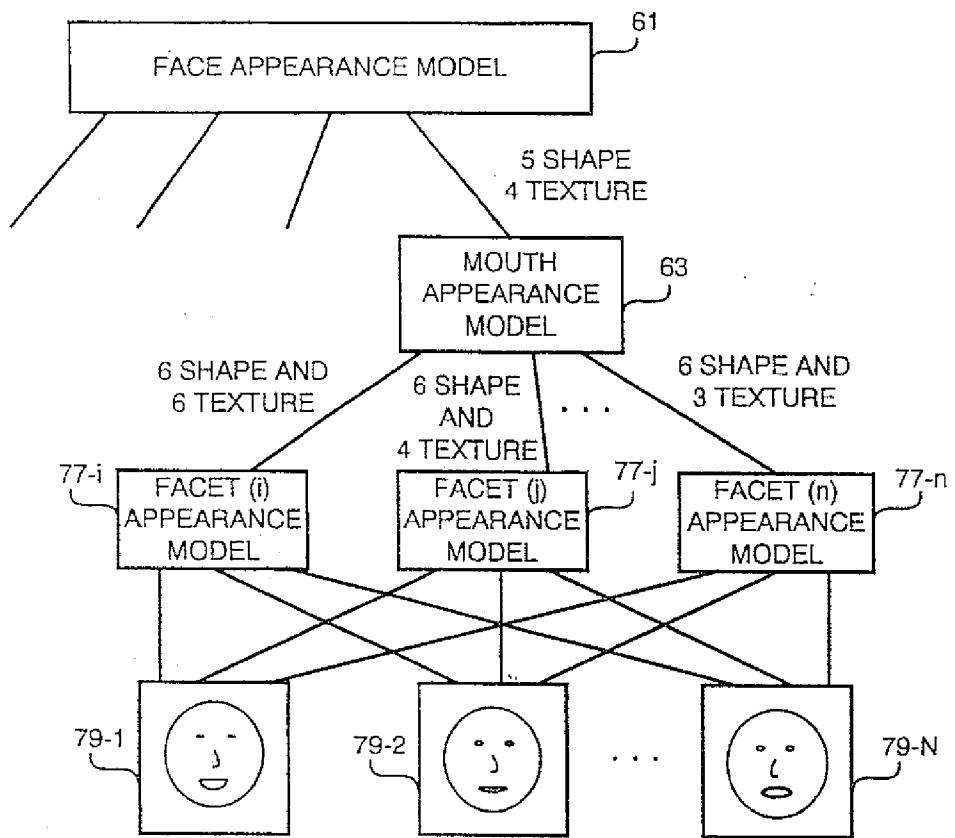


Fig. 10

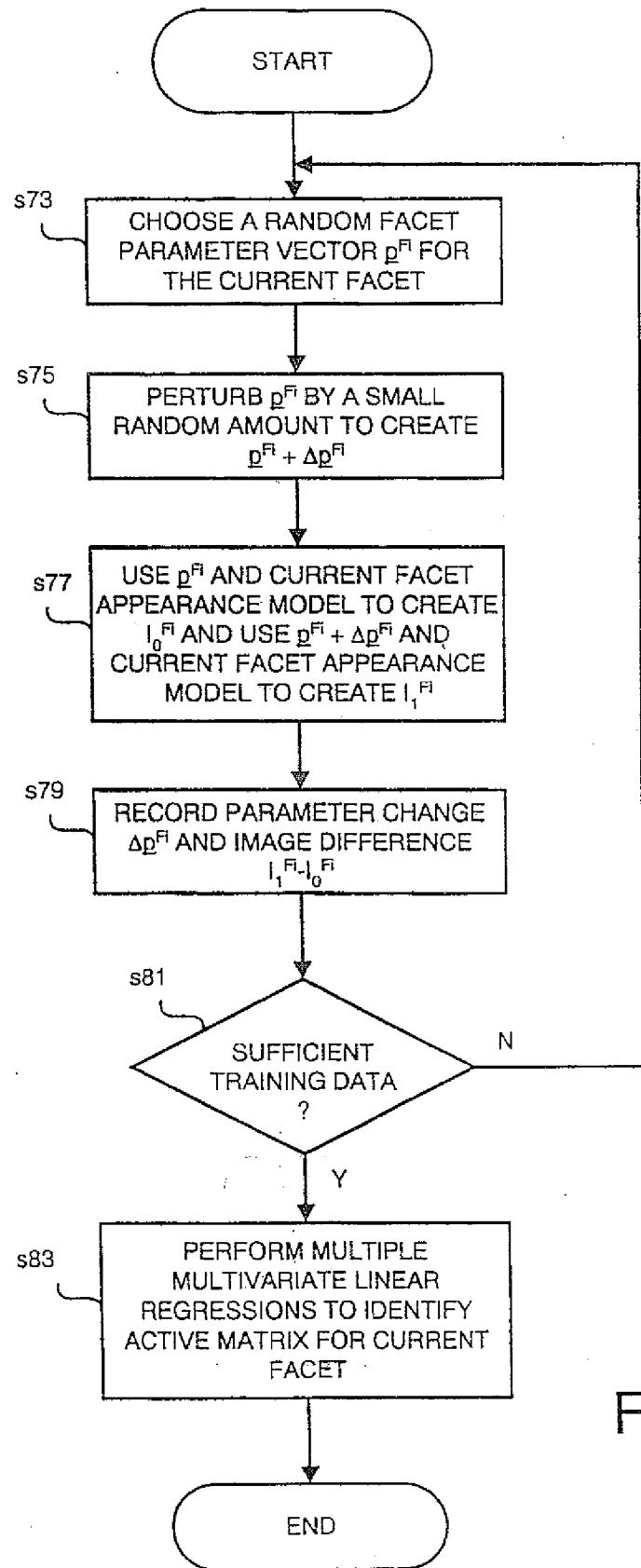


Fig. 11a

12/14

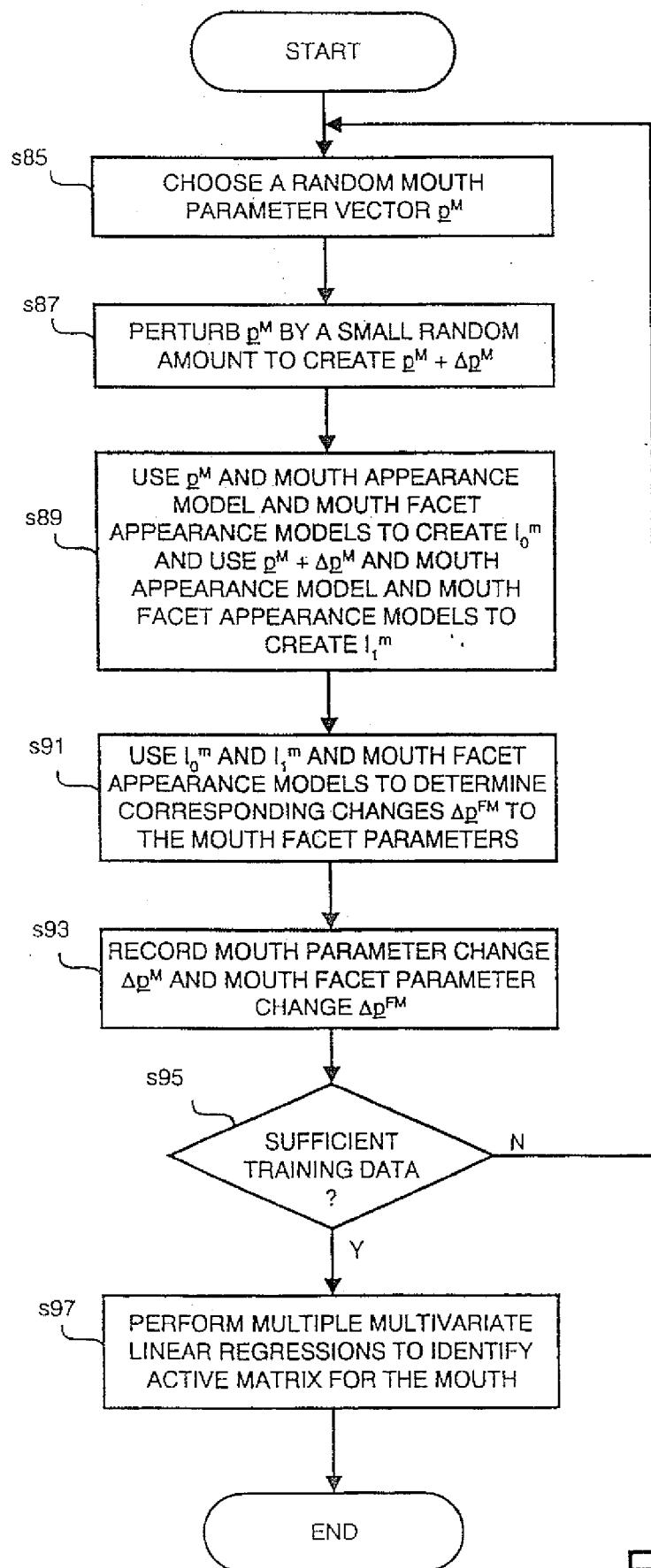


Fig. 11h

13/14

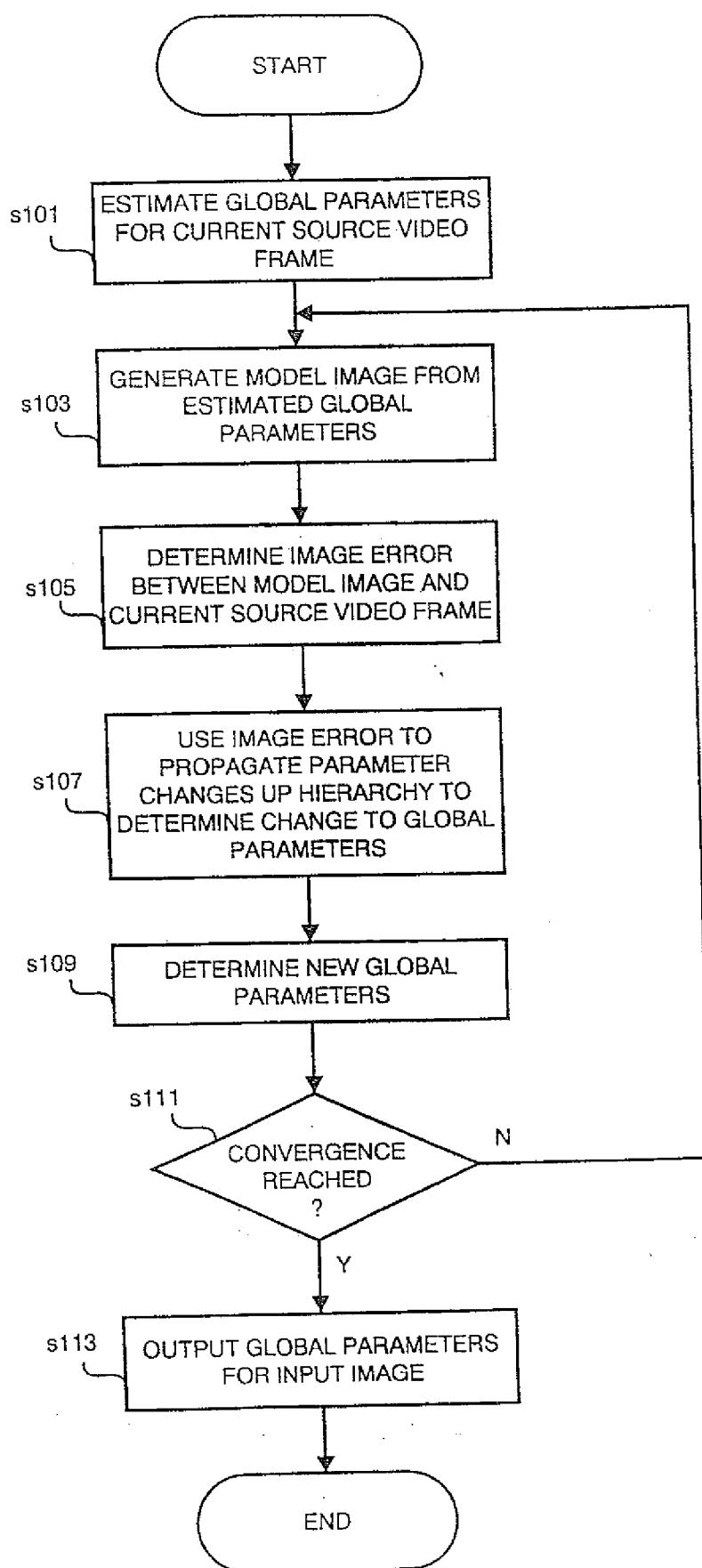


Fig. 12

14/14



Fig. 13a



Fig. 13b

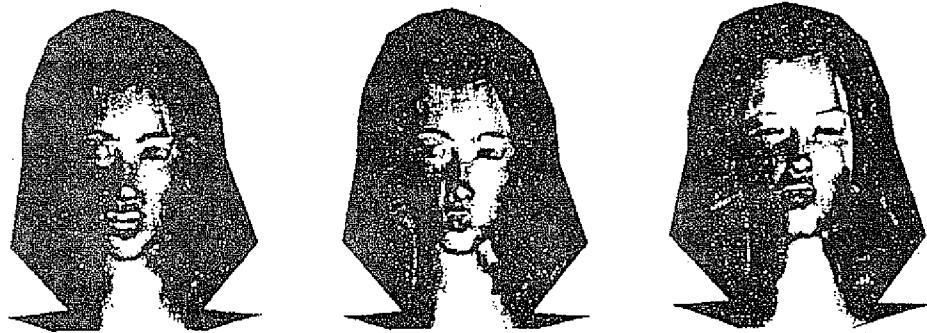


Fig. 13c



Fig. 13d



INTERNATIONAL SEARCH REPORT

Internat. Appl. No.

PCT/GB 00/04411

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 7 G06T17/00 G06K9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06K G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC, IBM-TDB, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>EDWARDS G. J ET AL: "ADVANCES IN ACTIVE APPEARANCE MODELS" KERKYRA, GREECE, SEPT. 20 - 27, 1999, LOS ALMITOS, CA: IEEE COMP. PRESS, US, vol. CONF. 7, 1999, pages 137-142, XP000980072 ISBN: 0-7695-0165-6 abstract page 137, left-hand column, paragraph 1 -right-hand column, paragraph 2 --- -/-</p>	1-32

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

° Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

26 April 2001

04/05/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl
 Fax: (+31-70) 340-3016

Authorized officer

König, W

INTERNATIONAL SEARCH REPORT

Internal Application No

PCT/GB 00/04411

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	COOTES T F ET AL: "ACTIVE SHAPE MODELS-THEIR TRAINING AND APPLICATION" COMPUTER VISION AND IMAGE UNDERSTANDING, ACADEMIC PRESS, US, vol. 61, no. 1, January 1995 (1995-01), pages 38-59, XP000978654 ISSN: 1077-3142 abstract page 38, left-hand column, paragraph 1 -page 39, left-hand column, paragraph 1	1-32

